

# How do genomes evolve?

Processes of genomic evolution

Dr Ben Skinner

Mammalian Molecular Genetics Group • Department of Pathology

bms41@cam.ac.uk • 01223 333 709

# Outline

Copy Number Variation (CNVs)

Mechanisms of variation

- Non-allelic homologous recombination

- Non-homologous end joining

Impact on evolutionary rearrangements

# Outline

## Copy Number Variation (CNVs)

### Mechanisms of variation

Non-allelic homologous recombination

Non-homologous end joining

### Impact on Evolutionary rearrangements

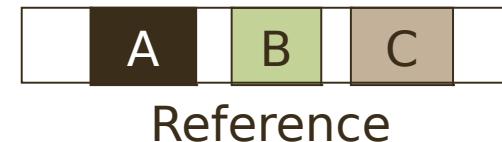
# Copy number variation

## Copy Number Variant (CNV)

>1kb segment of DNA existing in varying copy number between individuals in comparison with a reference genome

(not including insertions, deletions or transposable elements)

Range from 1kb to several Mb

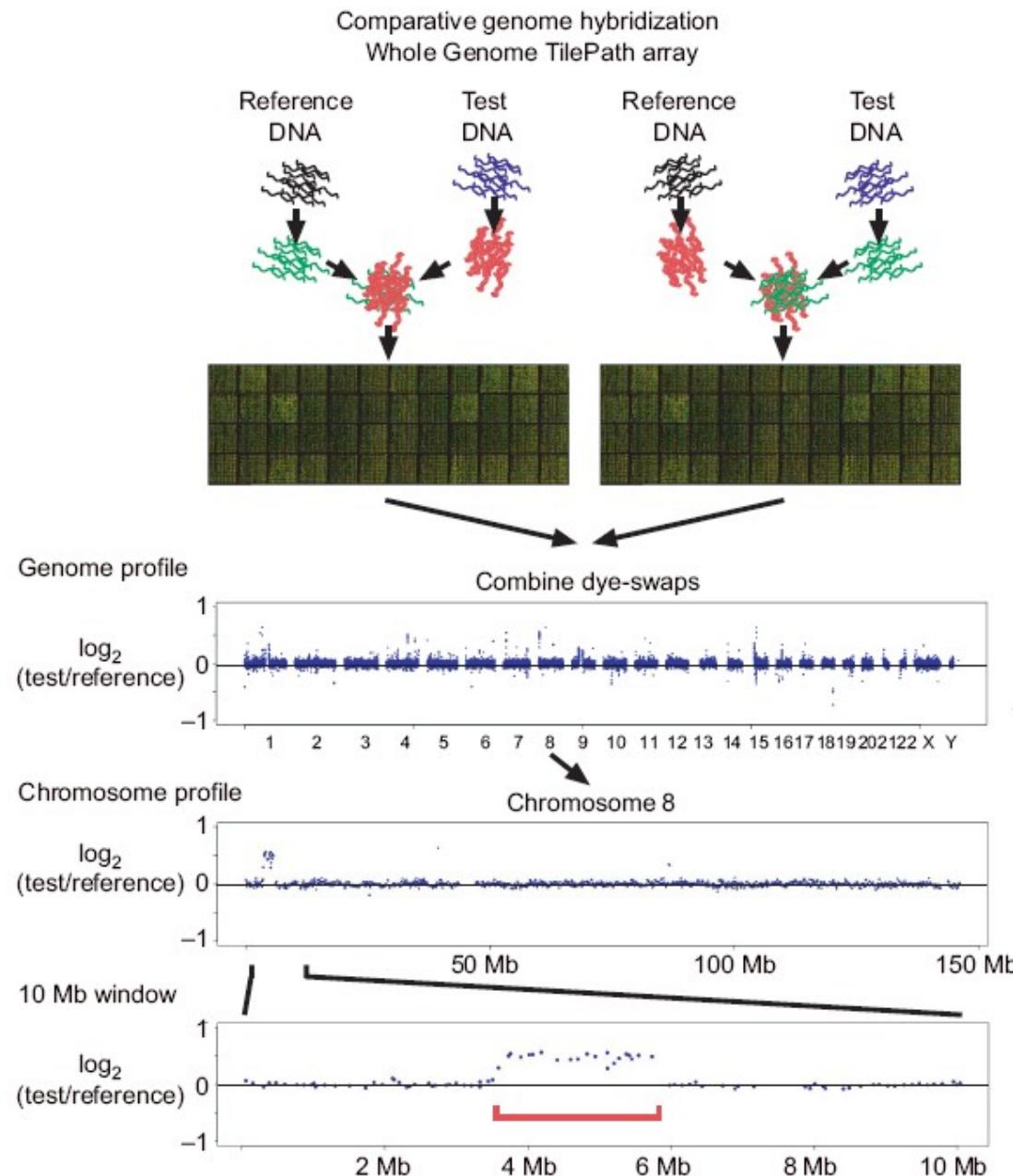


## Segmental duplication (SD)

Lower copy number  
Usually larger, less frequent.

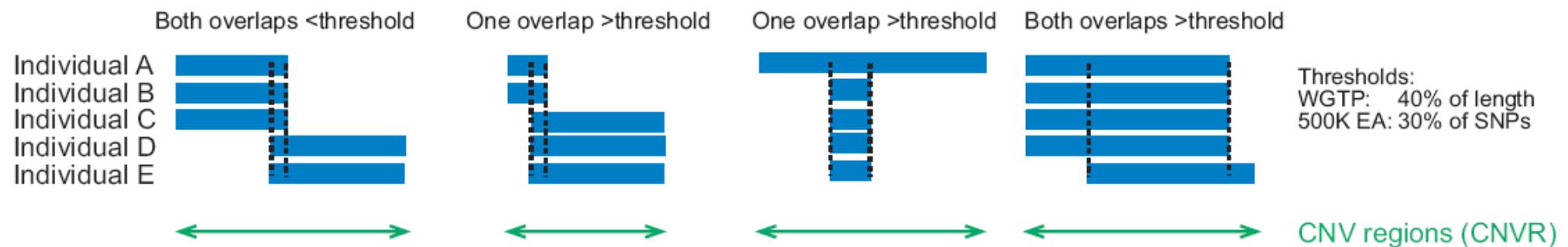
# Detecting CNVs

Microarray based comparative genomic hybridisation (Array-CGH)



# Copy number variation

CNVRs (copy number variable regions) are comprised of overlapping CNVs from different individuals



# CNVRs in 270 human genomes

International DNA and cell line collection derived from apparently healthy individuals:

30 parent offspring trios Yoruba Nigeria (YRI)

30 parent offspring trios European descent Utah (CEU)

45 unrelated Japanese Tokyo (JPT)

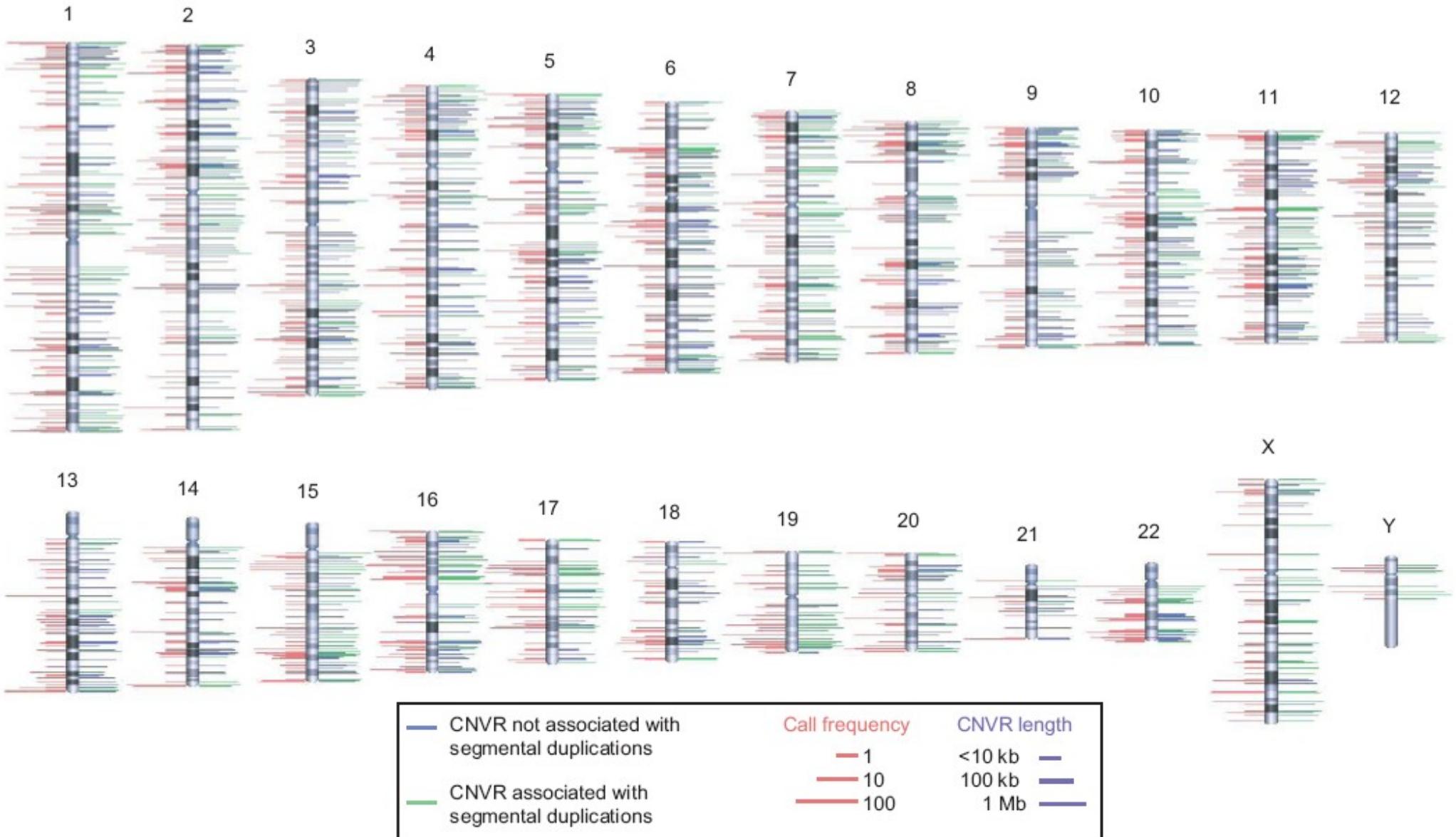
45 unrelated Han Chinese Beijing (CHB)

1447 CNVs in 980 CNVRs found covering 12% of genome (360Mb)

Median size 81kb

~50% of CNVRs in >1 individual

# CNVRs in 270 human genomes



CNVs are found across the genome – however, some regions are more prone to CNVs than others

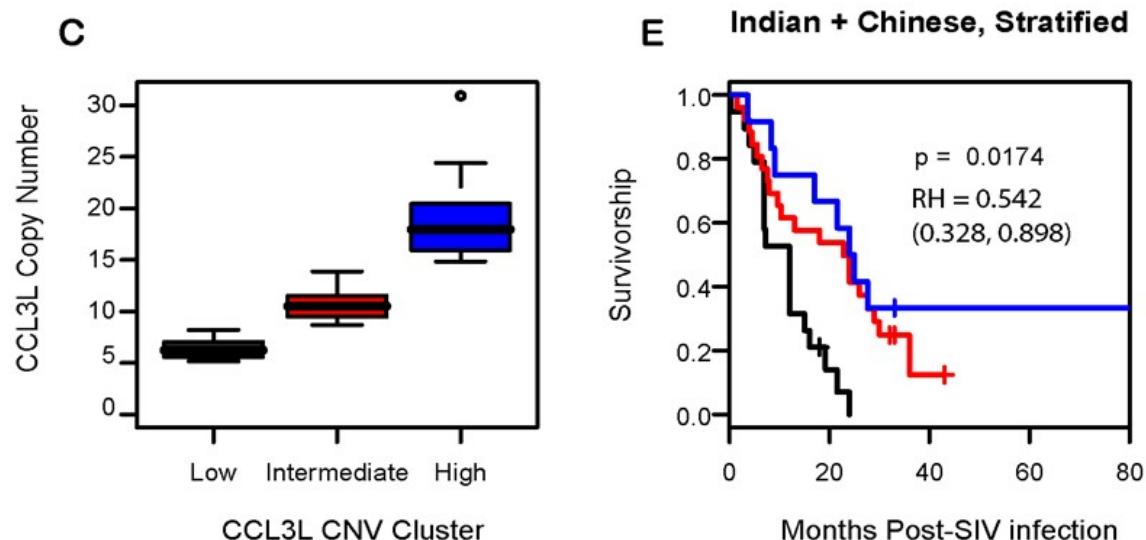
From *Nature* (2006)  
444: 444-54

# Copy number variation

CNVs are implicated in disease susceptibility and normal phenotypic variation

Examples:

Copy number of *CCL3-like* genes affects rate of progression to simian AIDS in Rhesus macaques



From *PLoS Genet* (2009)  
5(1): e1000346

Duplication of the  $\beta$  amyloid locus *APP* (amyloid precursor protein) causes autosomal dominant early-onset Alzheimer's disease

Effect on drug responses – targeted treatments

# Copy number variation

Selection in humans?

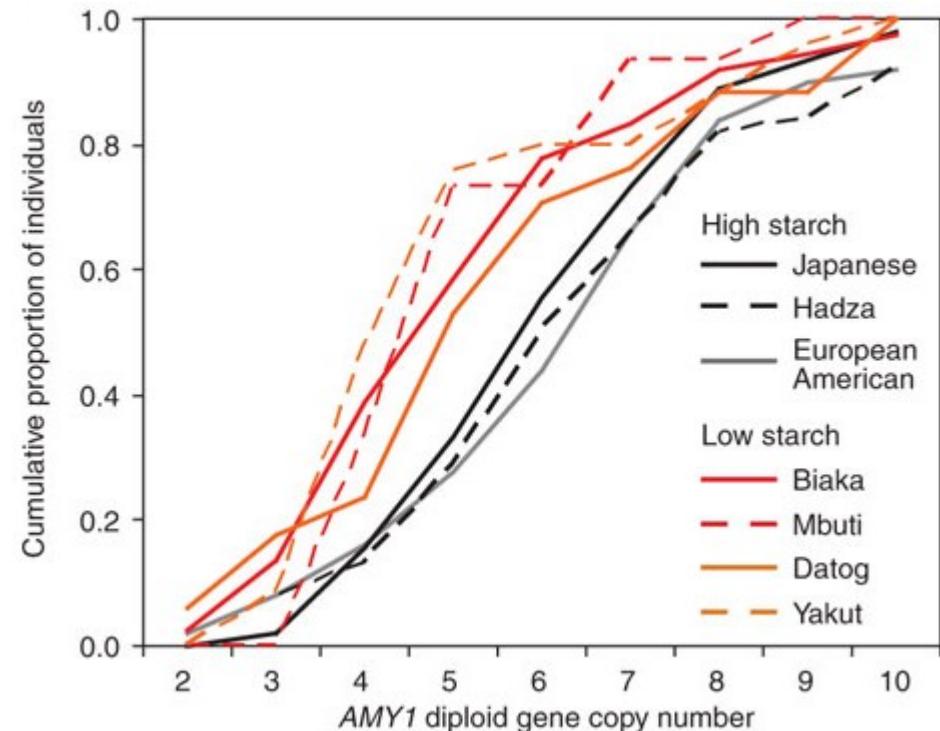
*AMY1* (salivary amylase)

Starch digestion. Higher copy number in populations with high starch diet.

*AQP7* (aquaporin)

Water transport across membranes. Several duplications of this gene specific to human lineage. Selection related to endurance running?

Supports the hypothesis that gene duplication is a major force of evolutionary change



From *Nat. Genet* (2007) 39: 1256-60

CNVs are candidate loci for speciation and disease related genes

# Copy number variation across species

CNVRs have been detected between species using the same approaches

So far, mainly in mammals e.g. human-chimp comparisons

Also a few birds e.g. chicken-duck, chicken zebra-finch

In each case, there is evidence for lineage specific CNVs

Patterns of CNVs also vary between lineages (numbers, sizes, chromosomal distribution)

The mechanisms that generate this variation?

# Outline

Copy Number Variation (CNVs)

Mechanisms of variation

- Non-allelic homologous recombination

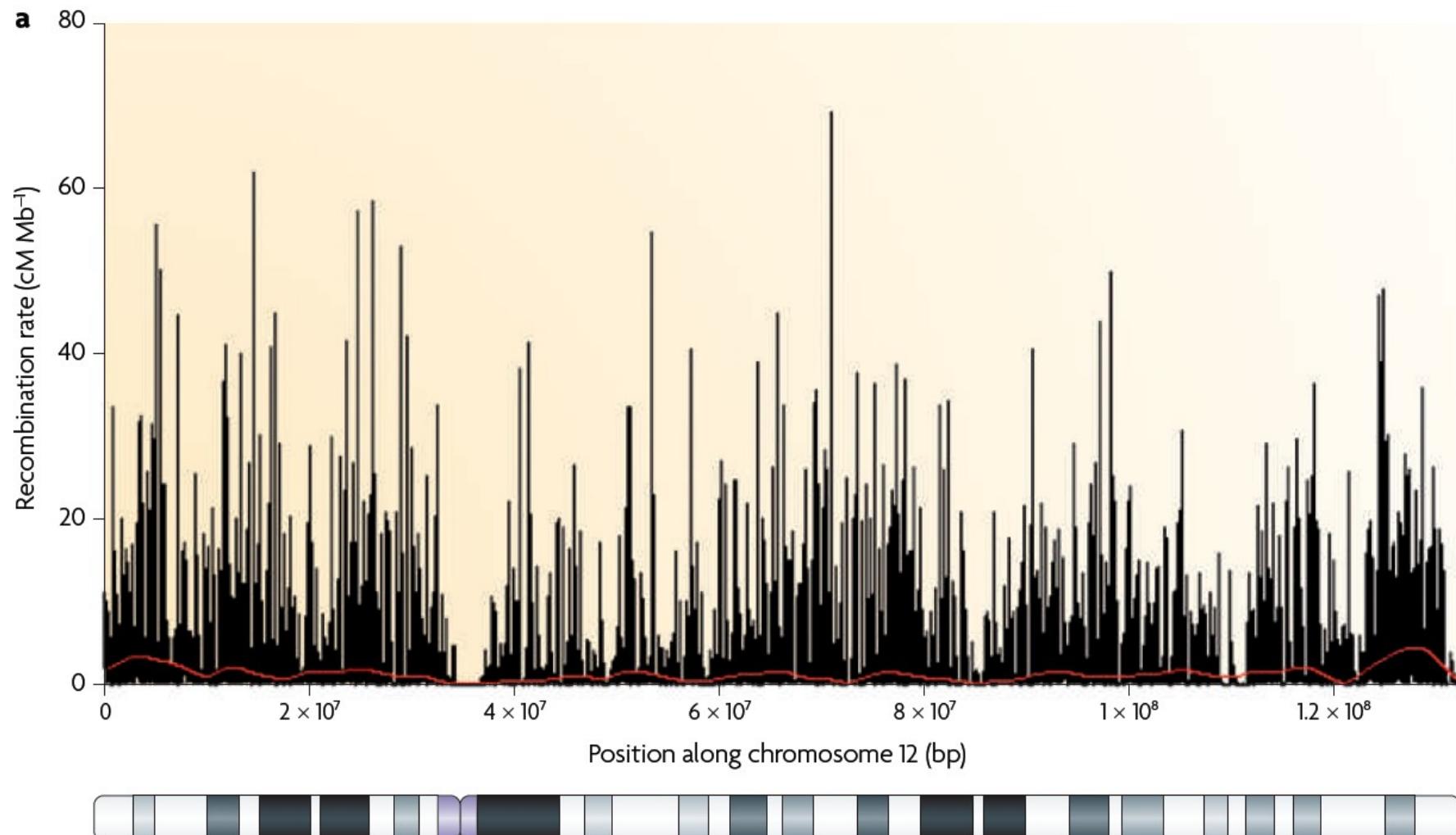
- Non-homologous end joining

Impact on evolutionary rearrangements

# Recombination rates vary across genomes

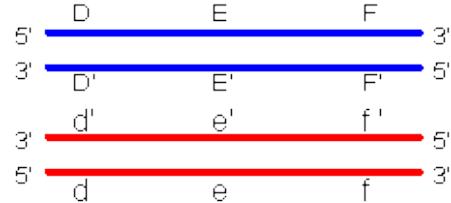
(and can vary substantially between species)

Hotspots of recombination across the genome



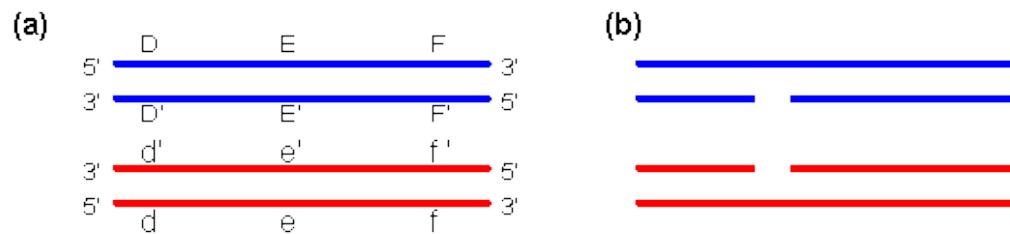
# Recombination

(a)



Pairing of homologous chromosomes in meiosis

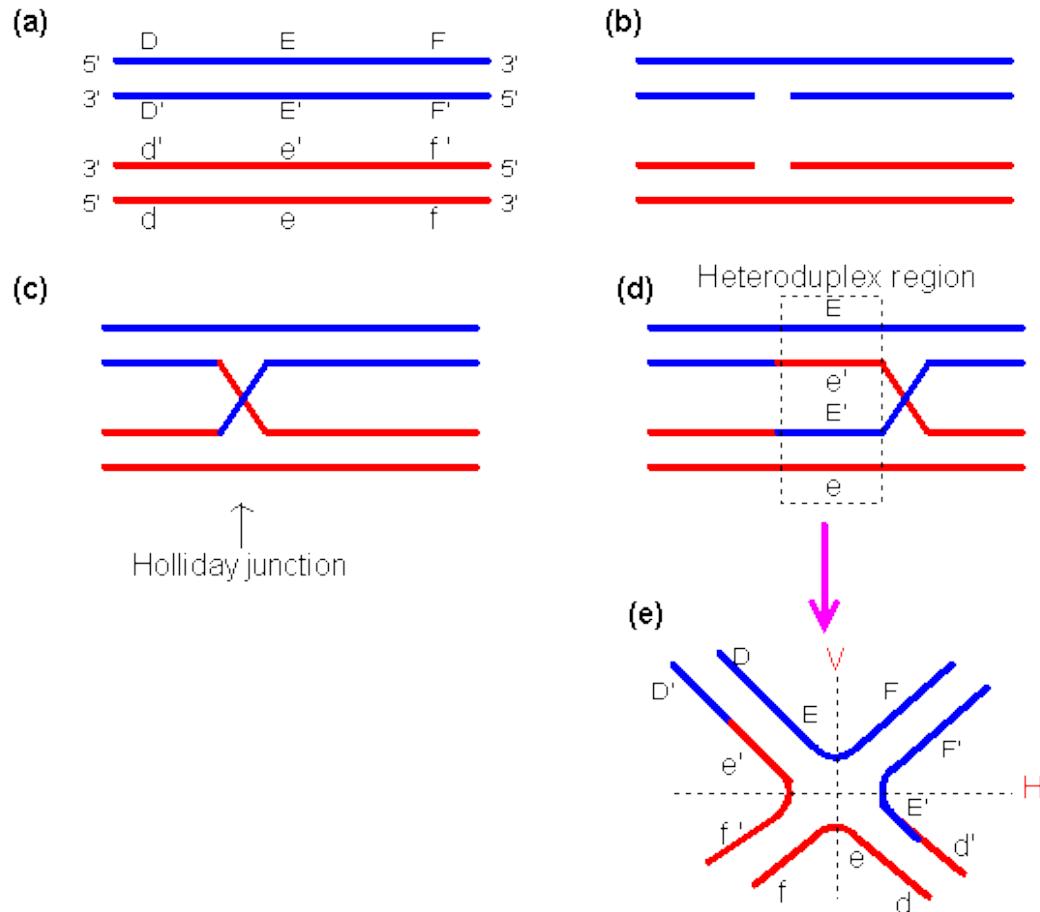
# Recombination



Pairing of homologous chromosomes in meiosis

Double strand breaks form

# Recombination

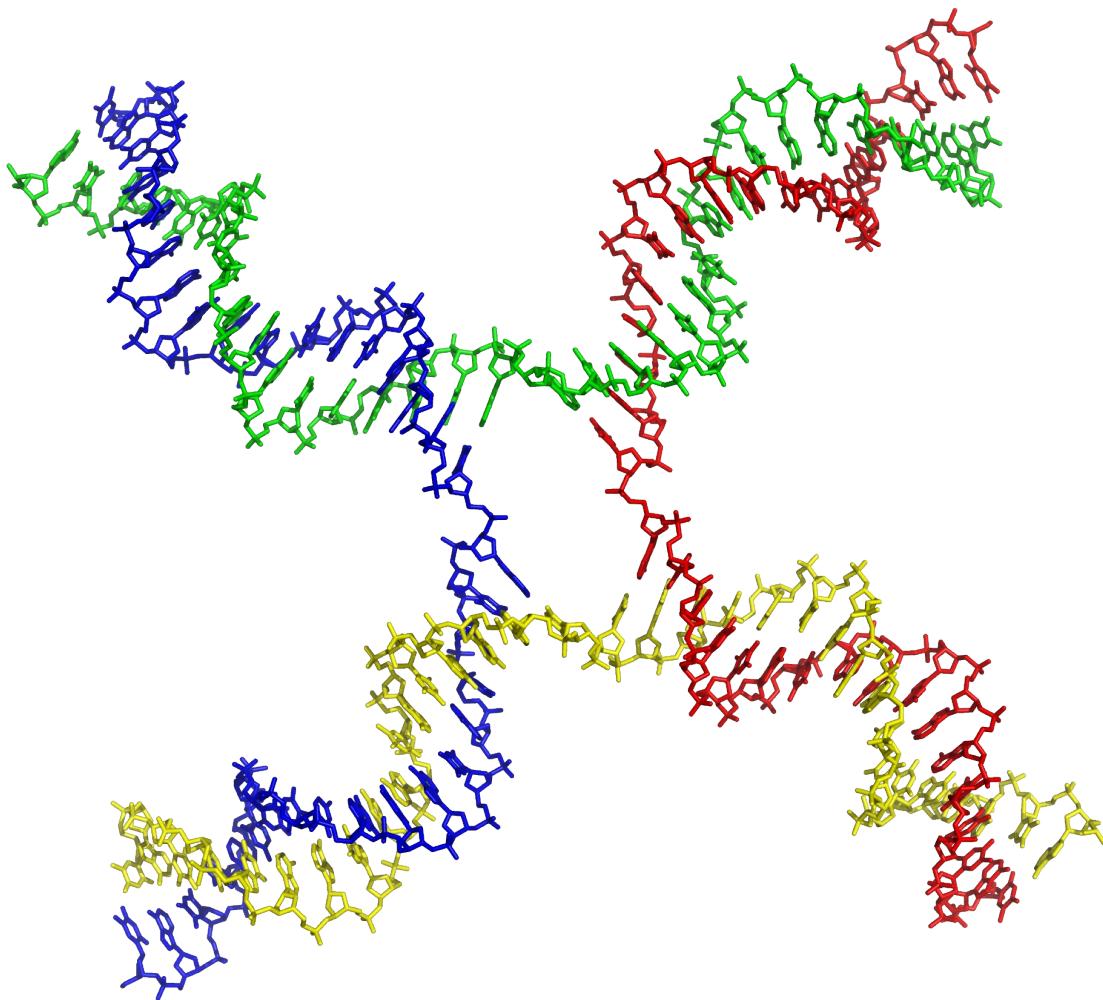


Pairing of homologous chromosomes in meiosis

Double strand breaks form

Cut strands invade and form Holliday junction

# Recombination

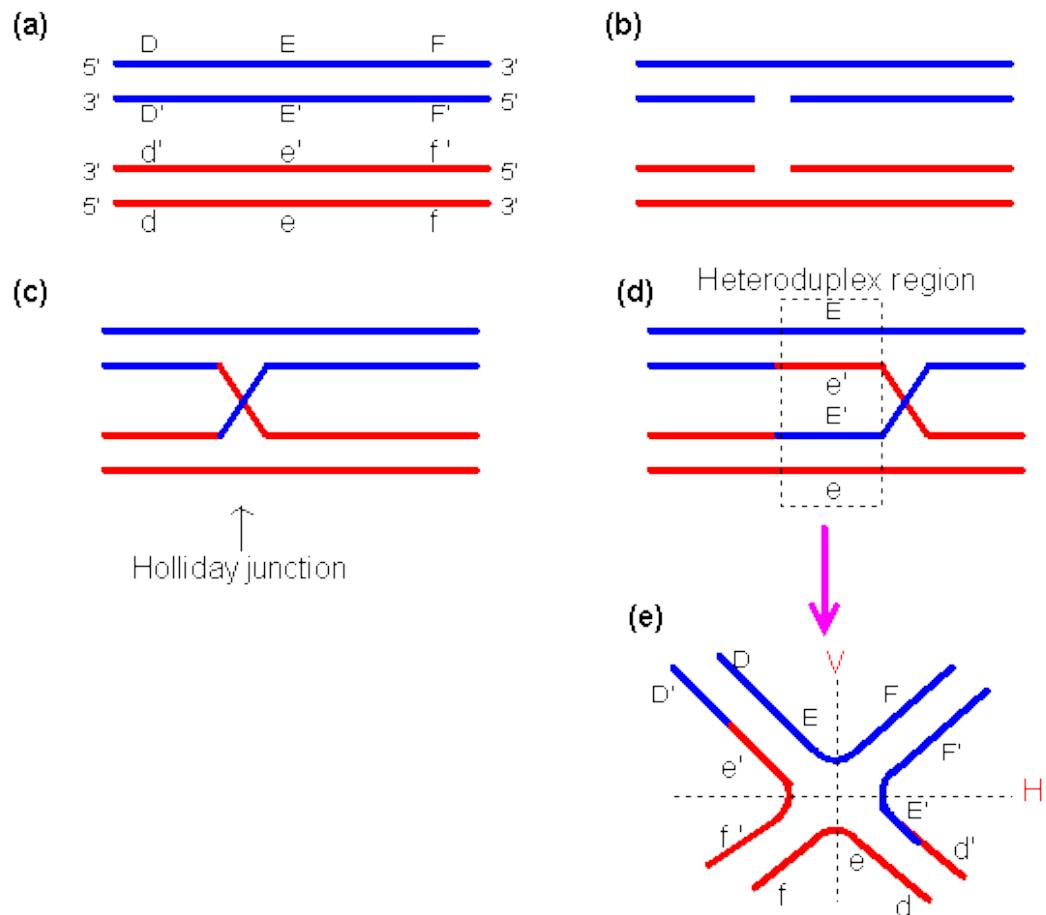


Pairing of homologous chromosomes in meiosis

Double strand breaks form

Cut strands invade and form Holliday junction

# Recombination

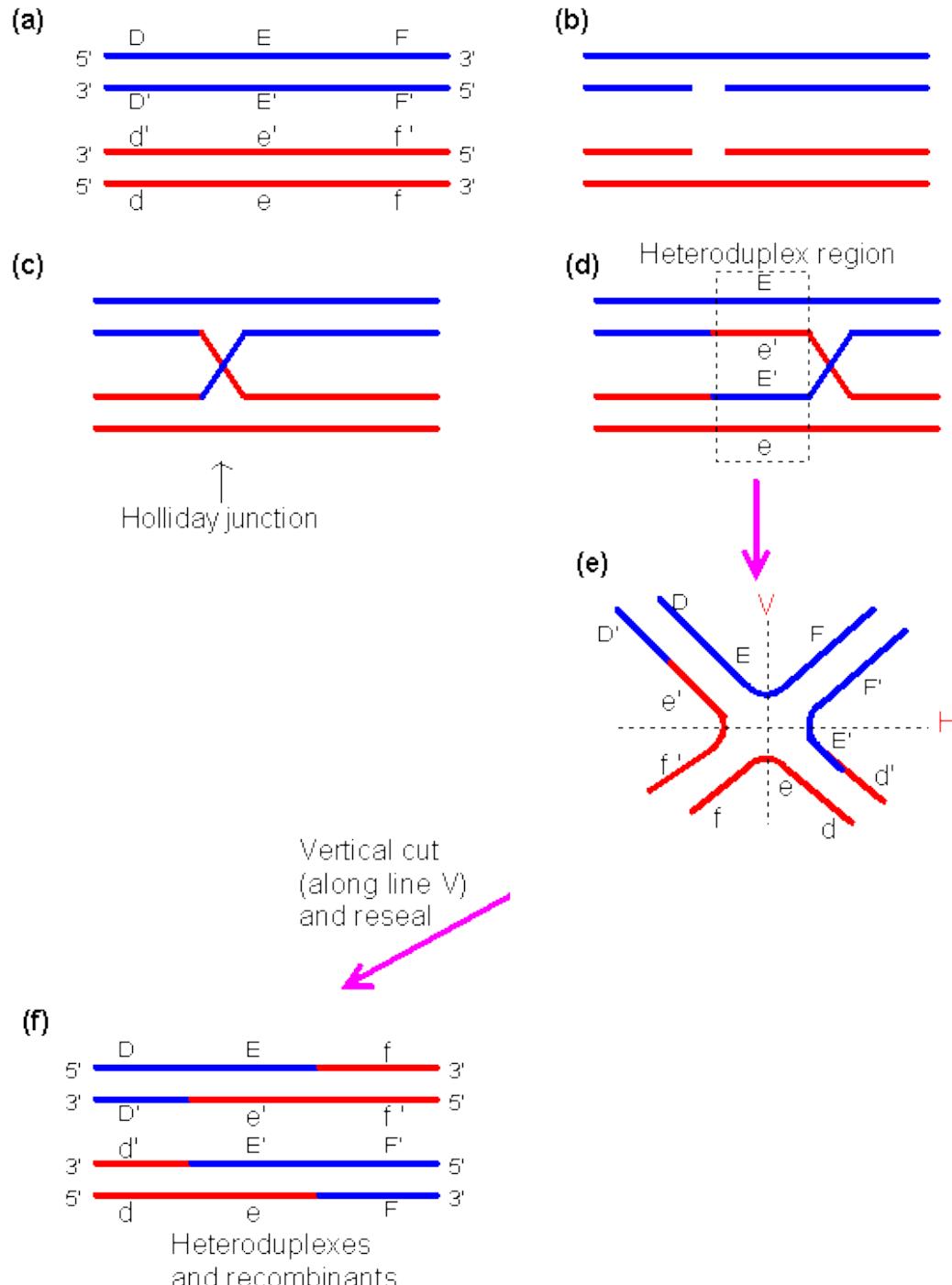


Pairing of homologous chromosomes in meiosis

Double strand breaks form

Cut strands invade and form Holliday junction

# Recombination



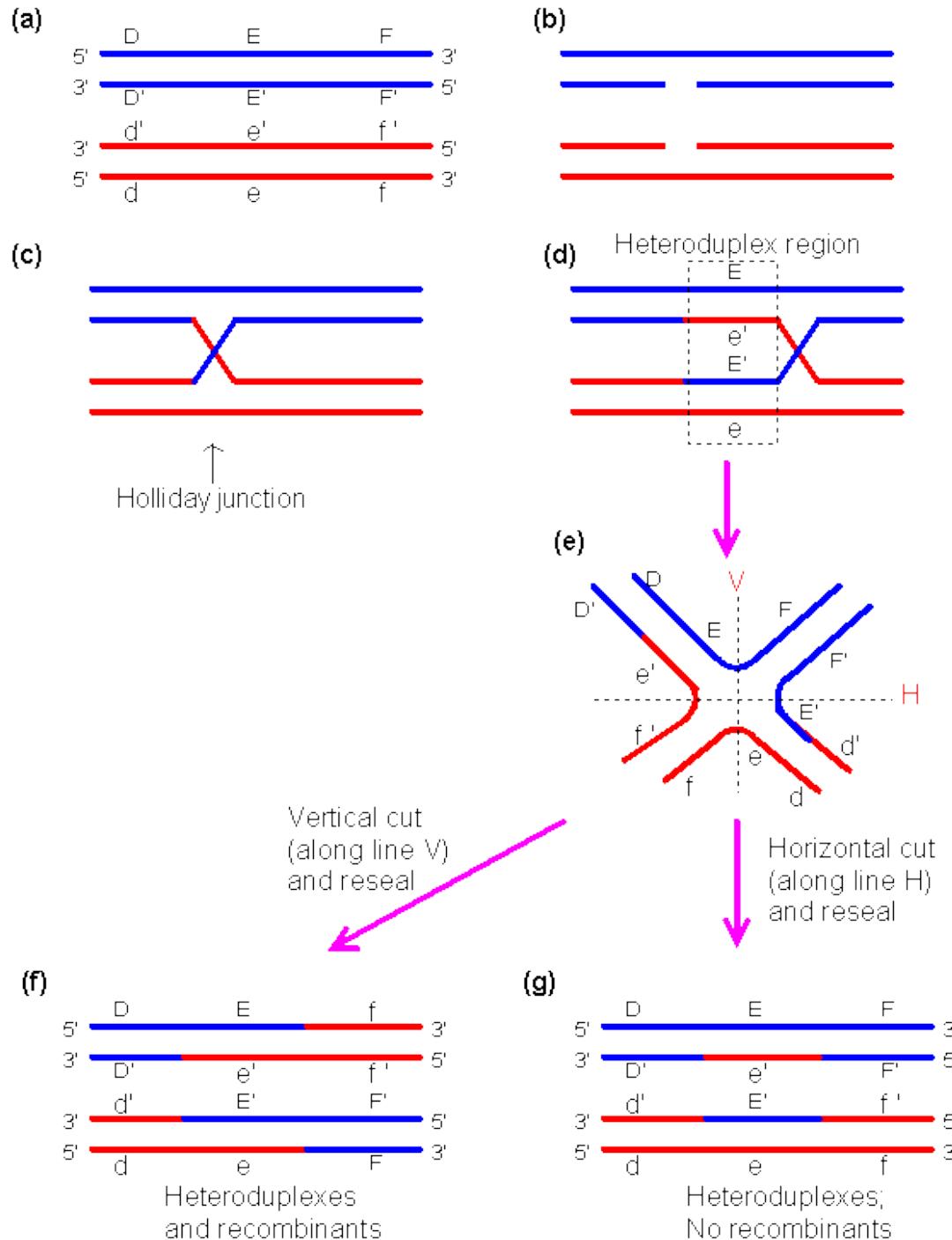
Pairing of homologous chromosomes in meiosis

Double strand breaks form

Cut strands invade and form Holliday junction

Resolution of crossover as heteroduplexes with or without recombinants depending on which strands are cut

# Recombination



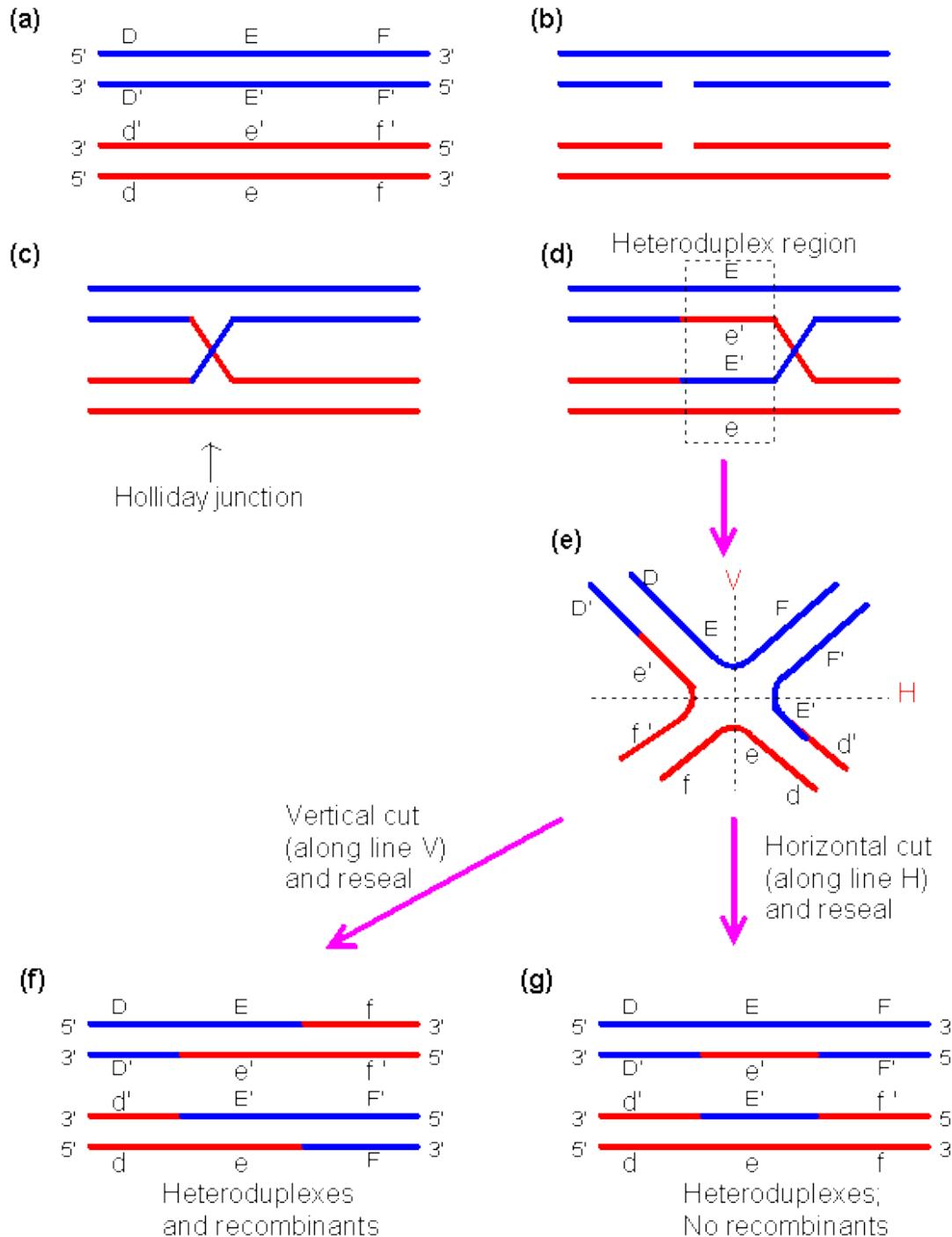
Pairing of homologous chromosomes in meiosis

Double strand breaks form

Cut strands invade and form Holliday junction

Resolution of crossover as heteroduplexes with or without recombinants depending on which strands are cut

# Recombination



Pairing of homologous chromosomes in meiosis

Double strand breaks form

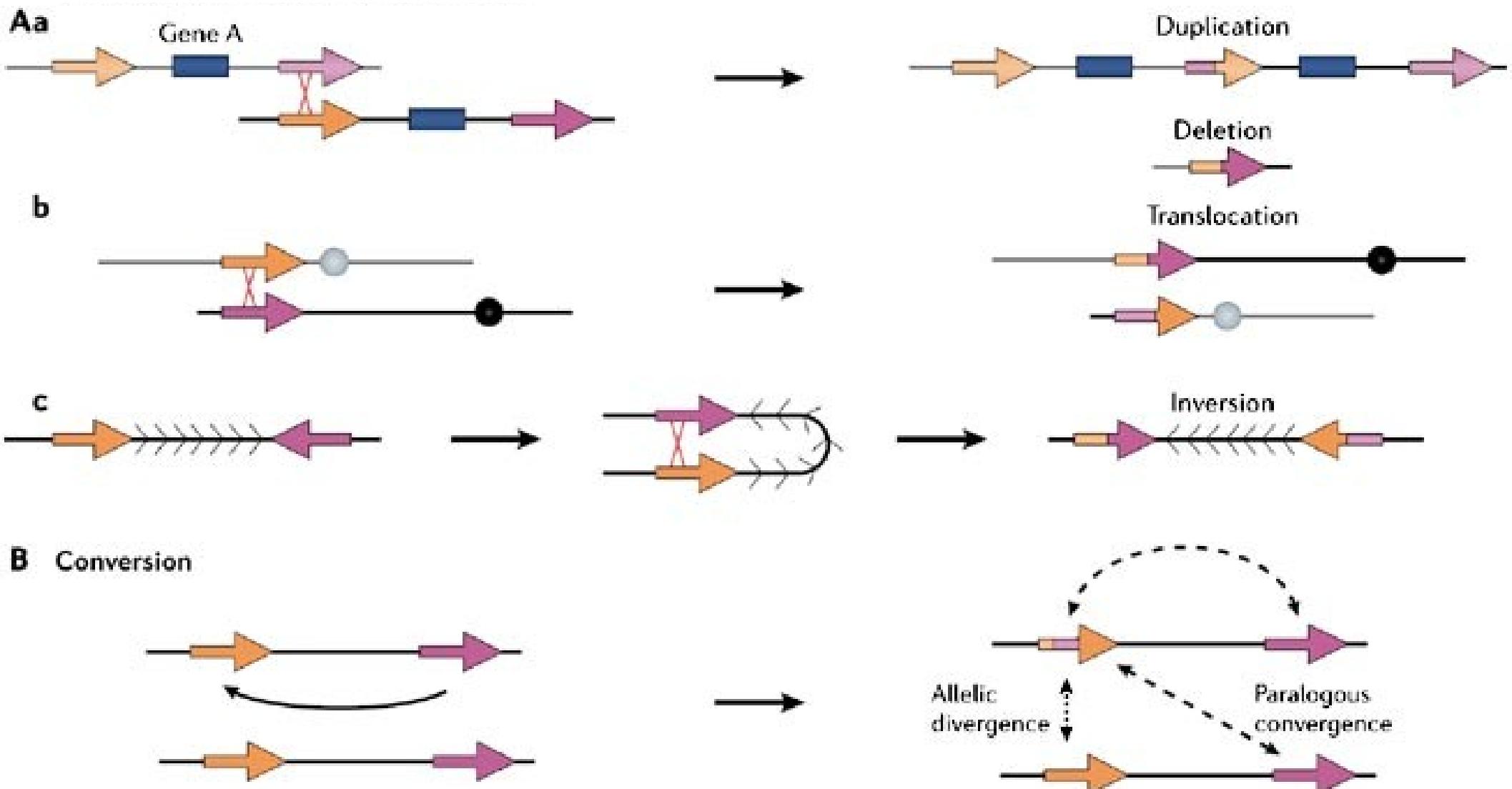
Cut strands invade and form Holliday junction

Resolution of crossover as heteroduplexes with or without recombinants depending on which strands are cut

Repair of mismatches in heteroduplex regions can lead to gene conversion

# Non-allelic homologous recombination (NAHR)

Recombination between non-allelic sequences (i.e. multicopy sequences)



# Non-homologous end joining (NHEJ)

Means of repairing double strand breaks in DNA

Breakpoint ends are directly ligated – no need for homologous template

## D-NHEJ

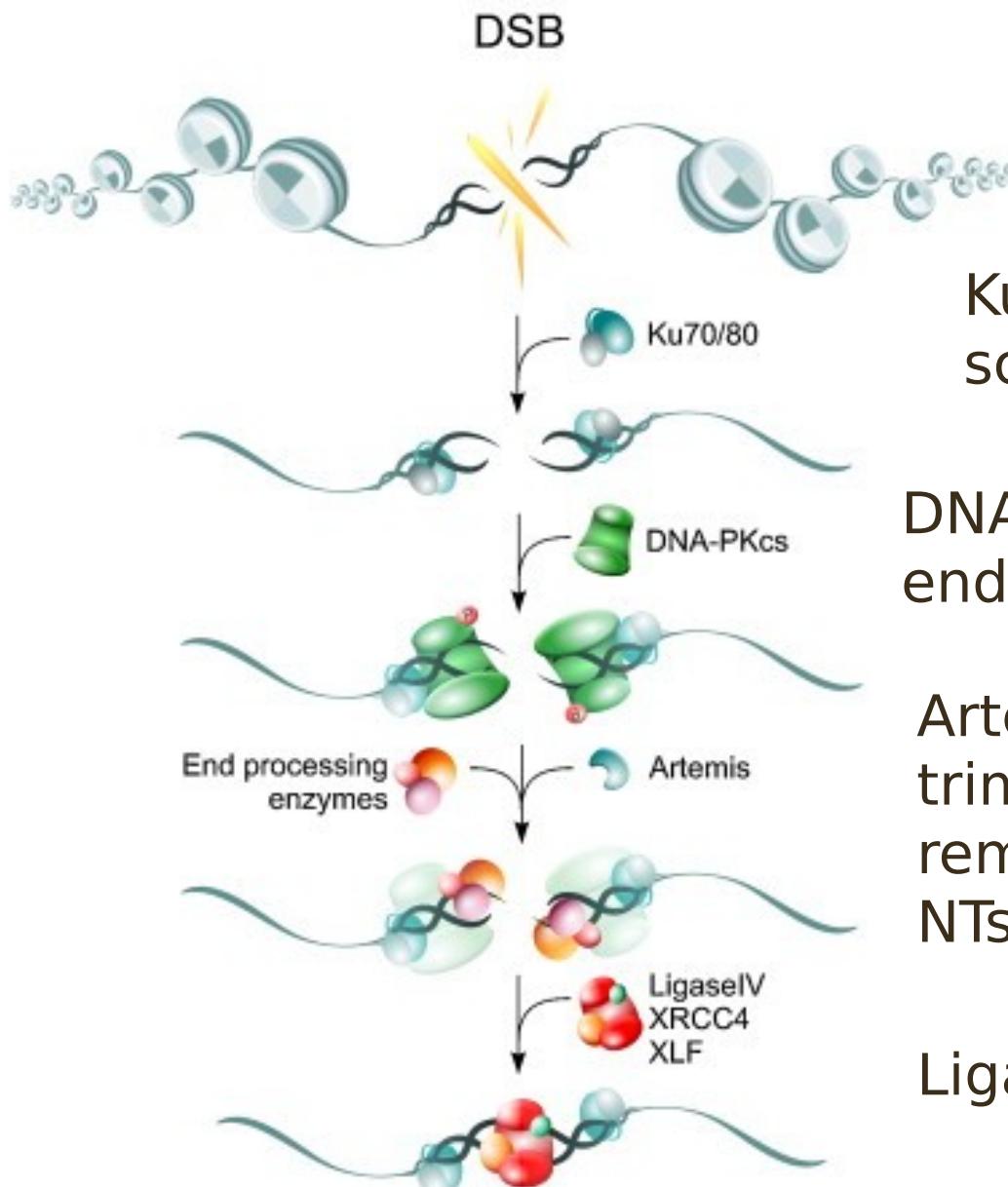
restores genomic integrity without ensuring sequence restoration

## B-NHEJ

alternative pathway of end joining operating with slower kinetics and mainly as backup to D-NHEJ

# Non-homologous end joining (NHEJ)

## D-NHEJ



Ku70/80 dimer binds DSB - scaffold

DNA-PKcs allows access to free ends

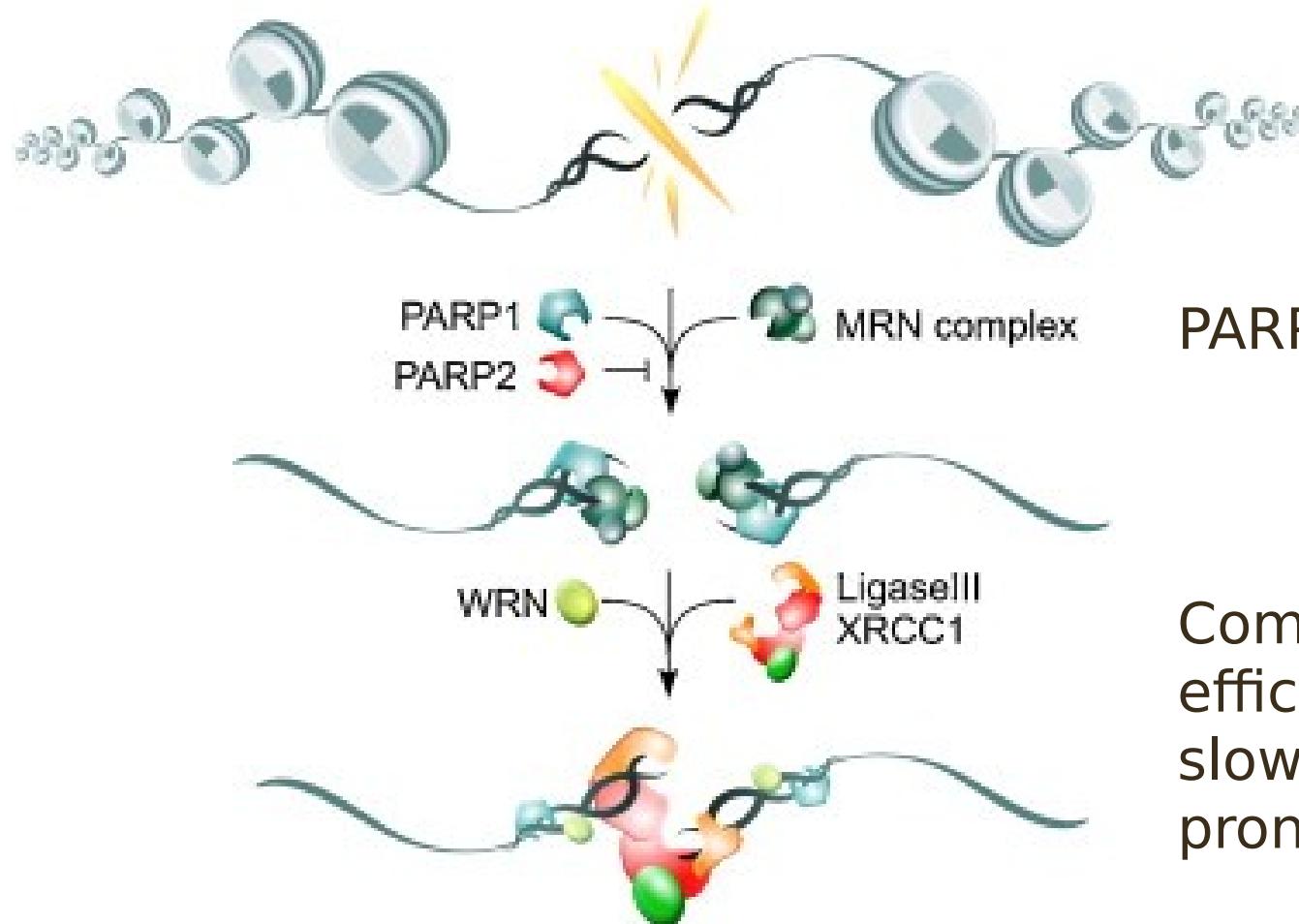
Artemis nuclease may aid trimming of ends. End processing removes damaged / mismatched NTs

Ligase IV complex joins free ends

# Non-homologous end joining (NHEJ)

## B-NHEJ

DSB



PARP similar to Ku - scaffold

Complex proteins less efficient than in D-NHEJ – slower and more error-prone

e.g. human tumors with non-functional DNA-PKcs

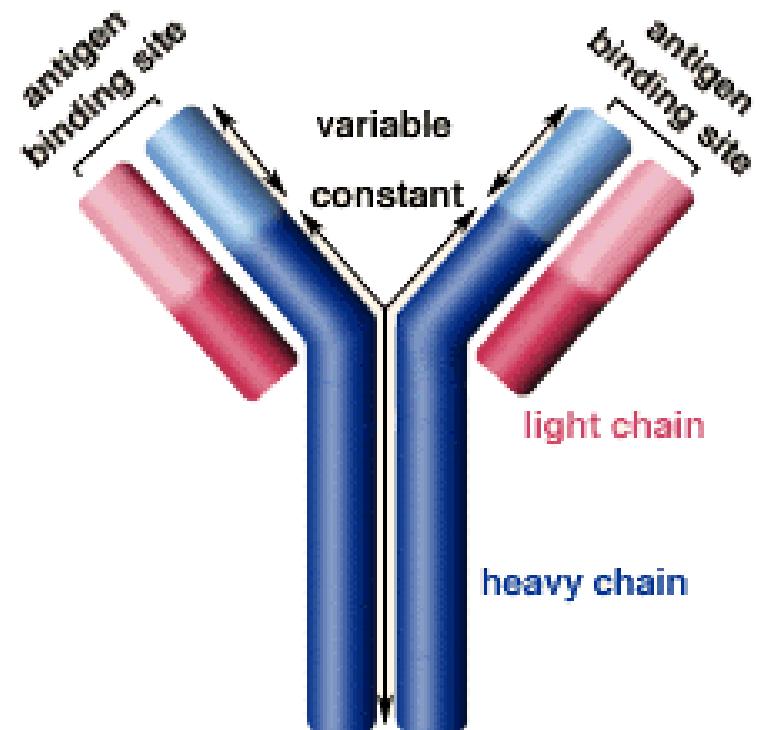
# Non-homologous end joining (NHEJ)

Critical role in V(D)J recombination  
in antibody generation

Variable (V), diversity (D), and  
joining (J) regions of heavy chain  
locus

Together, these form the variable  
region of the B- or T-cell receptor  
genes

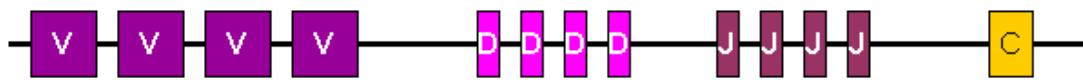
Error-prone repair by NHEJ is  
beneficial - maximizes diversity in  
coding sequence



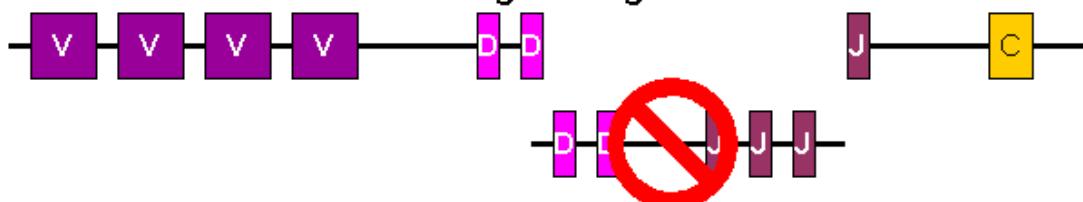
Patients with mutations in NHEJ genes cannot produce  
functional B or T cells - severe combined immunodeficiency  
(SCID)

# V(D)J Recombination

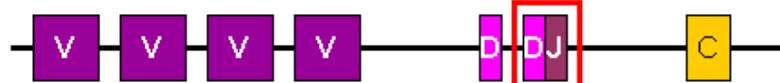
Genes in heavy chain locus



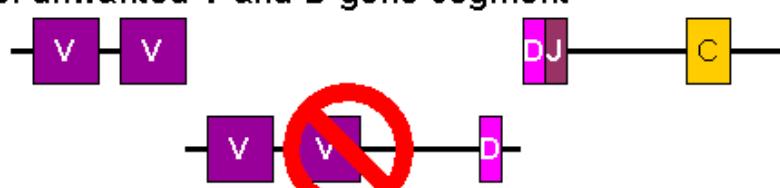
Removal of unwanted D and J gene segment



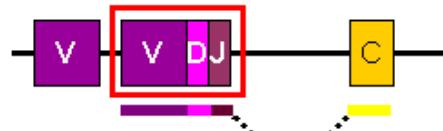
Recombination of D and J exons – DJ recombination



Removal of unwanted V and D gene segment



Recombination of V and DJ exons – VDJ recombination



Antibody transcript will also include constant domain gene

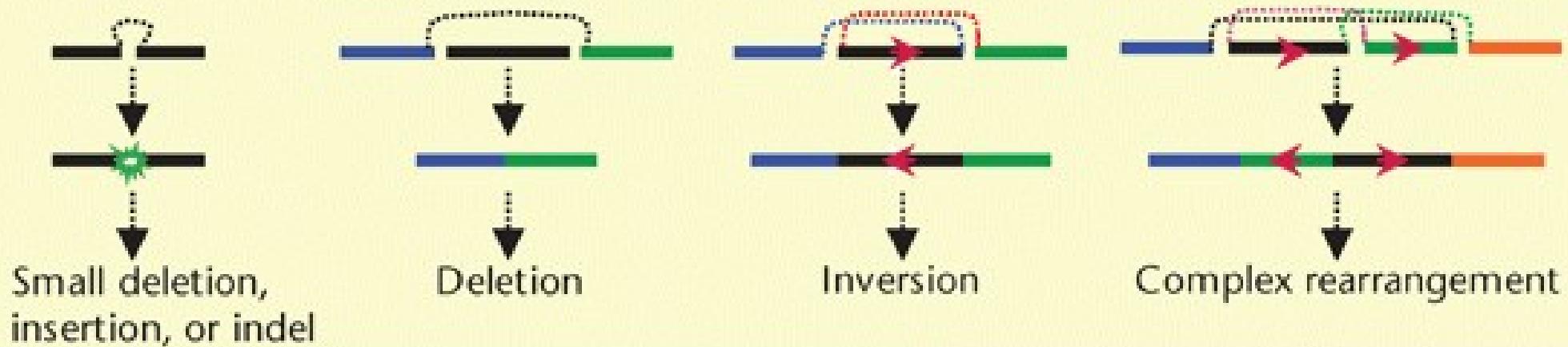
DNA cleaved at recombination signal by RAG1/RAG2 nuclease

DNA opened and joined by Artemis nuclease via NHEJ

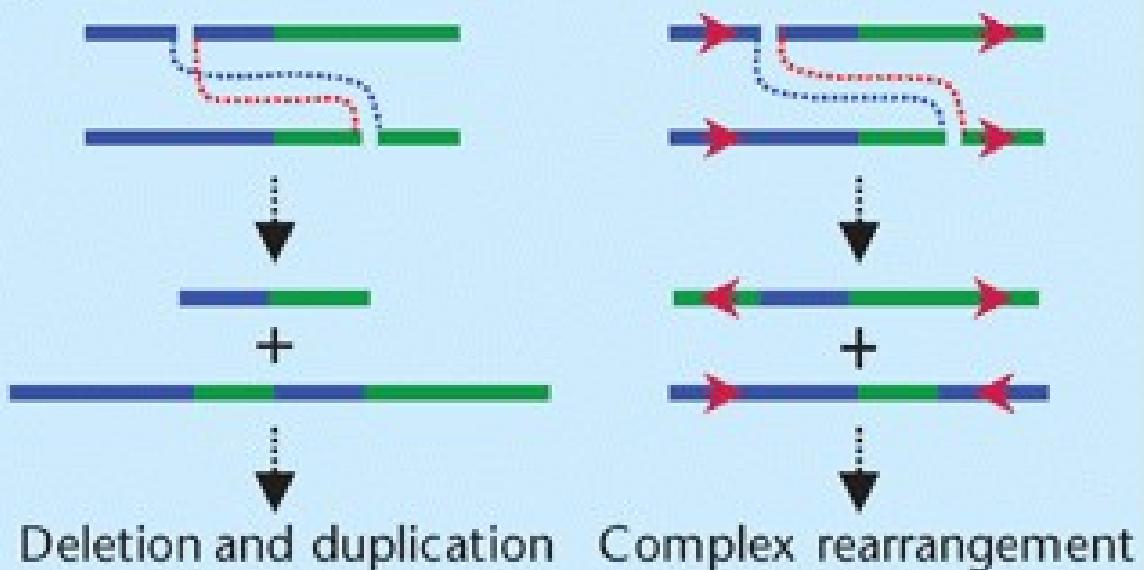
Generation of antibody diversity

# Non-homologous end joining (NHEJ)

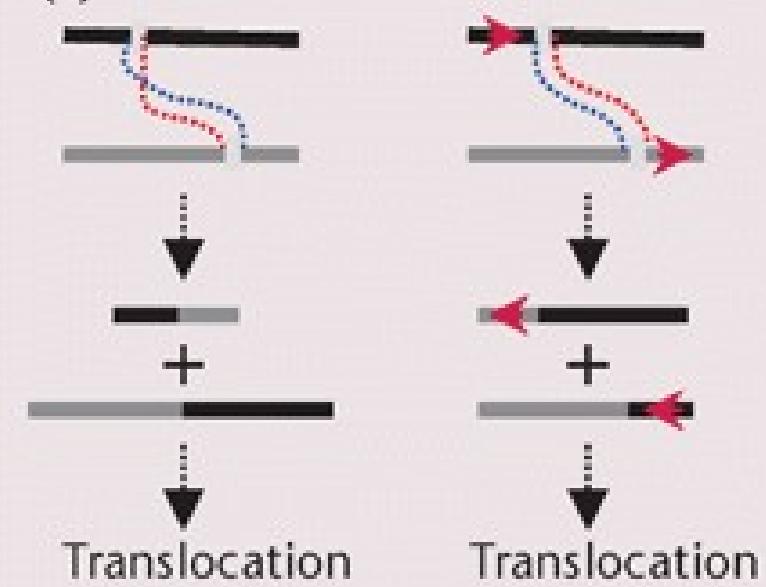
## (a) Intrachromosomal



## (b) IHC



## (c) INHC



# NAHR versus NHEJ

	NAHR	NHEJ
Breakpoints	Clustered	Scattered
Architectural Features	Low copy repeats (e.g CNVs)	Alu, LINE elements
Cis-acting stimulating sequence?	Transposons, minisatellites	Triplet repeats, telomeric repeats
Sequence features at recombinant junction	Gene conversion	Added bases at junctions

# Outline

Copy Number Variation (CNVs)

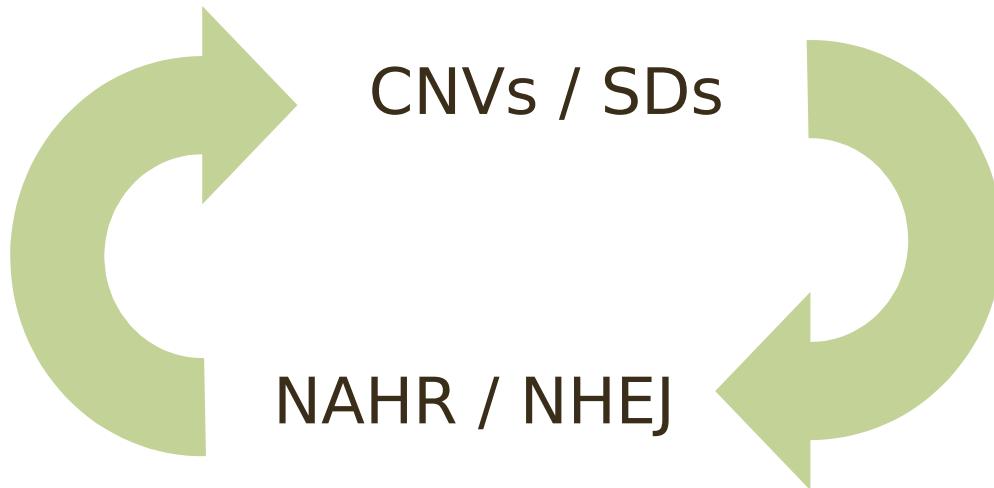
Mechanisms of variation

Non-allelic homologous recombination

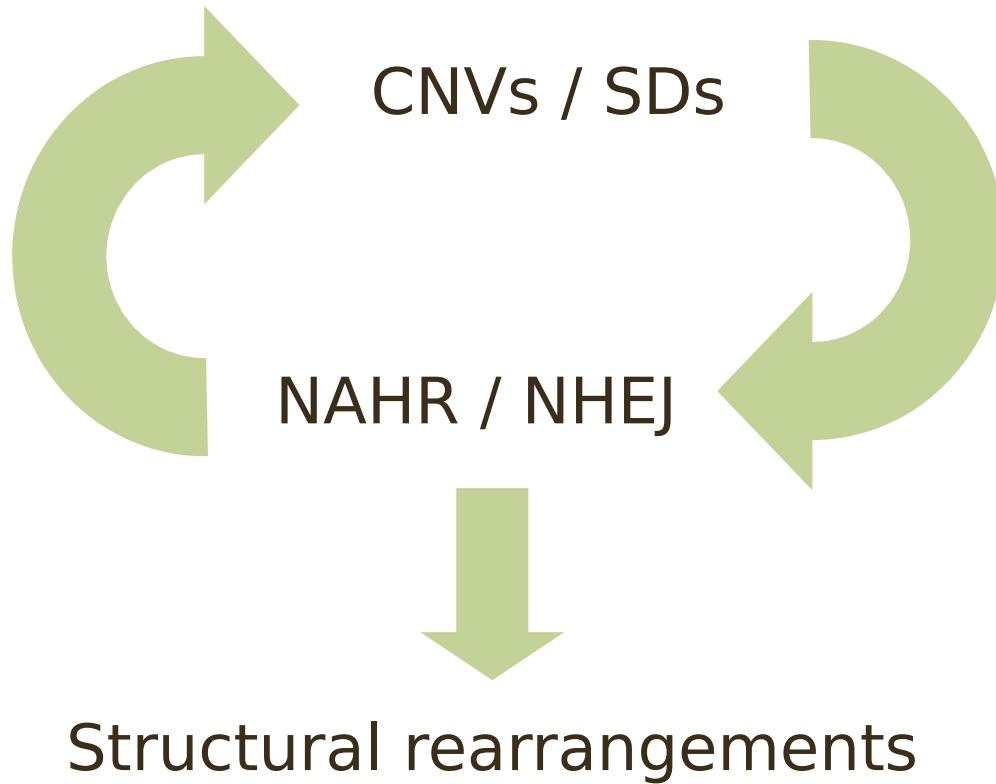
Non-homologous end joining

Impact on evolutionary rearrangements

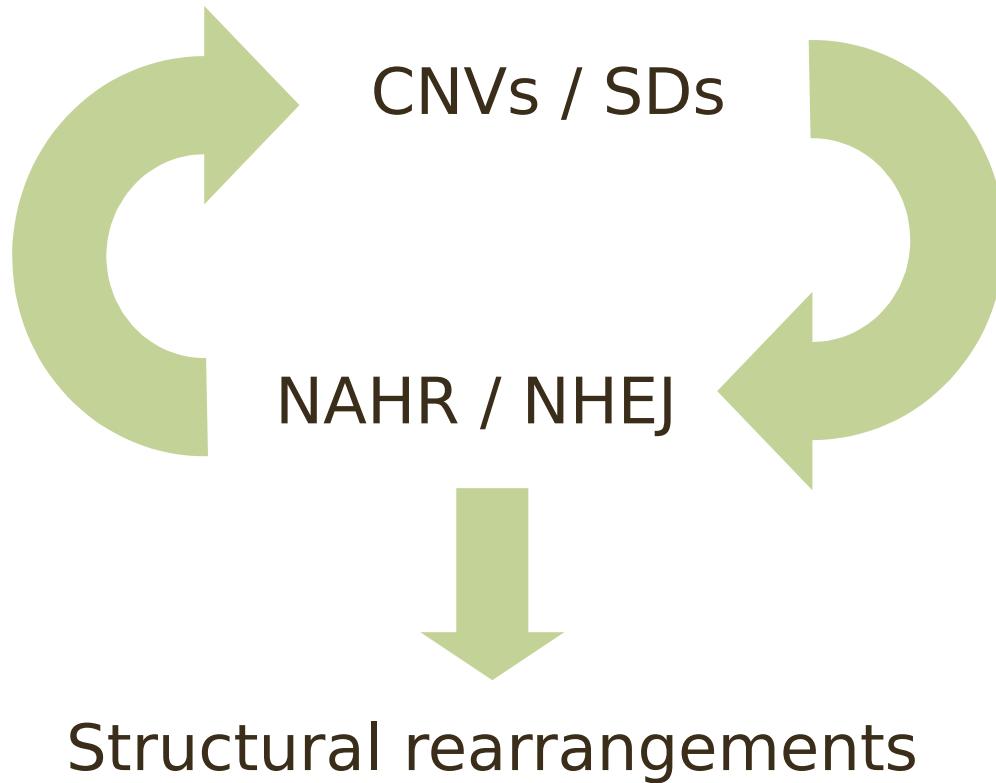
# Recombination drives CNV



# Recombination drives CNV



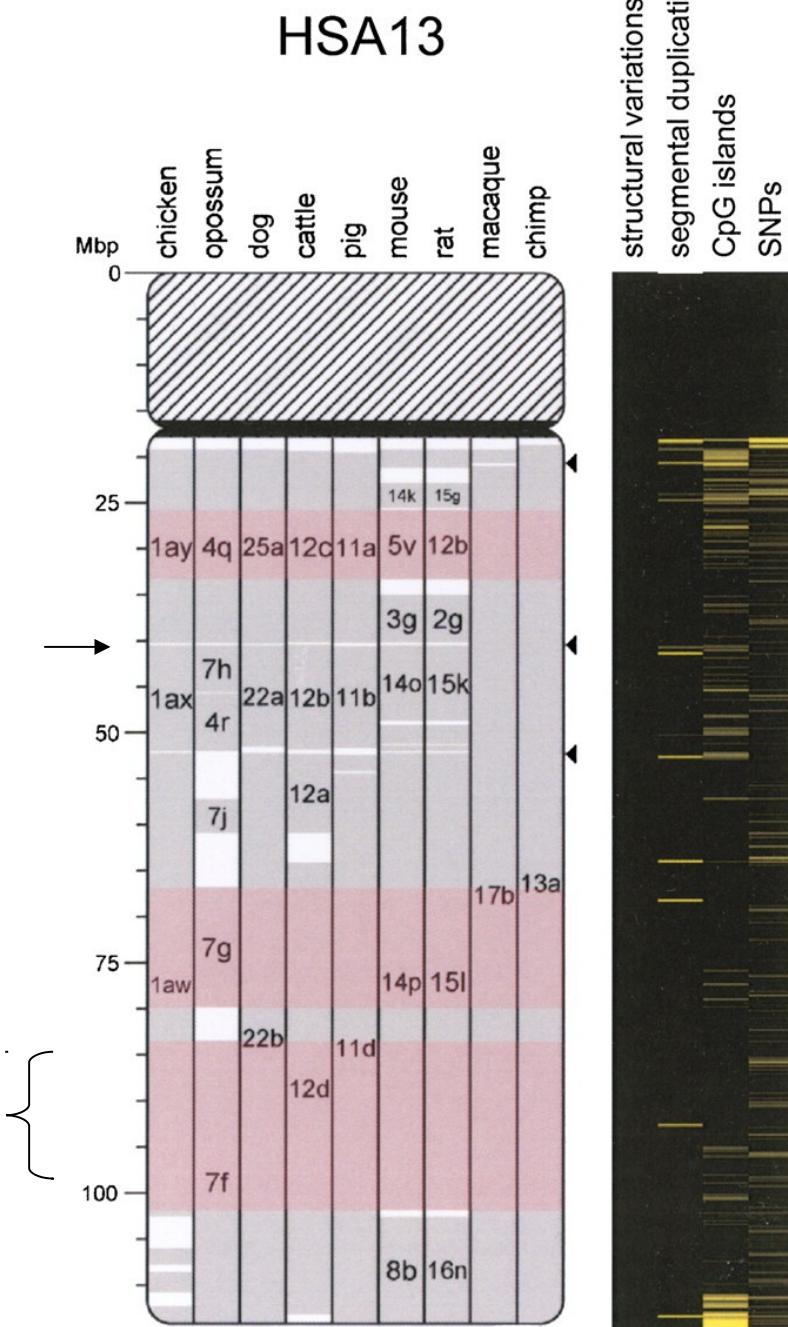
# Recombination drives CNV



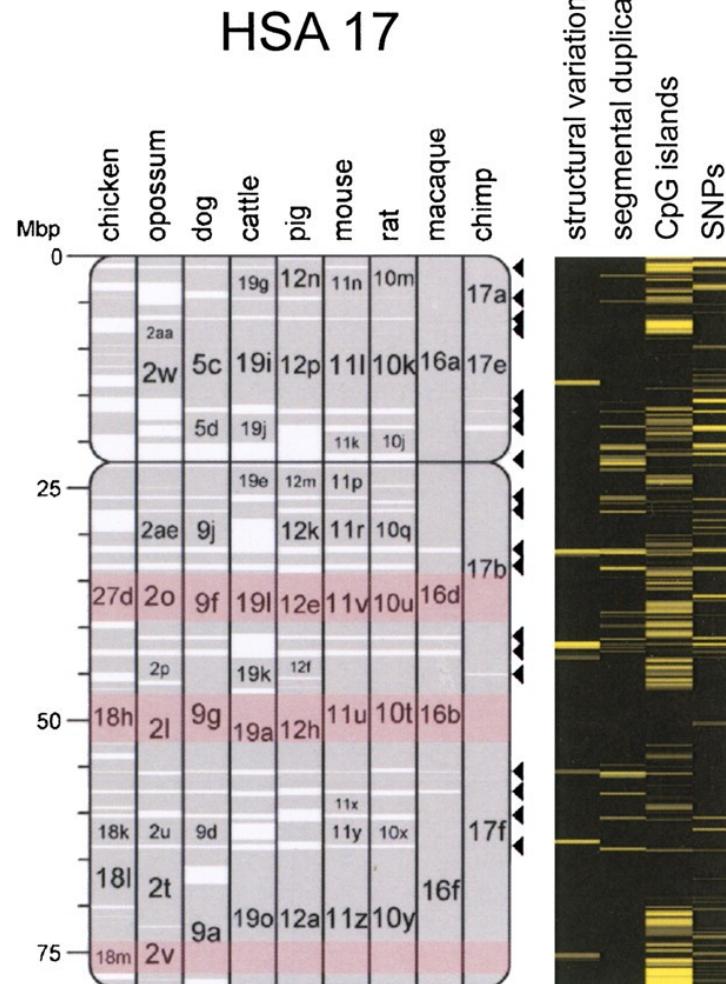
Prediction: Sites of CNV are linked with sites of evolutionary breakpoints

# Evolutionary Breakpoint Regions (EBRs)

A



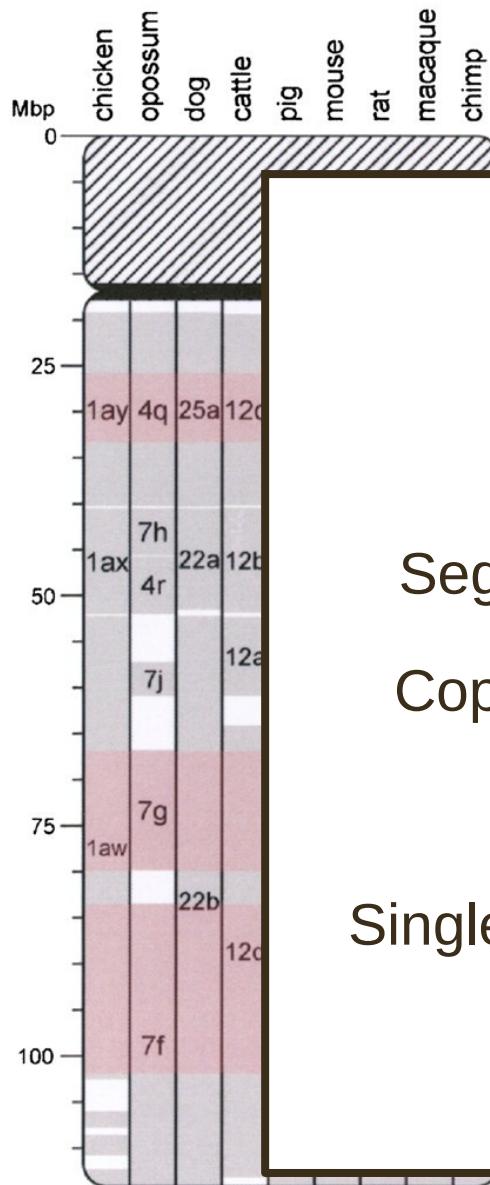
B



# Evolutionary Breakpoint Regions (EBRs)

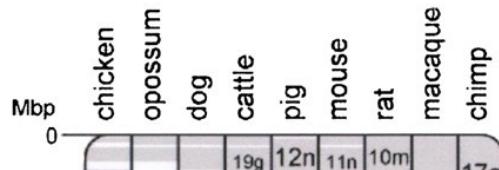
A

HSA13



B

HSA 17



EBRs are enriched for:

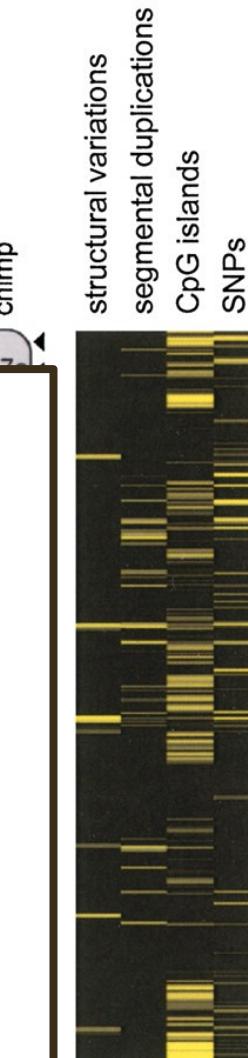
Structural variants

Segmental duplications (SDs)

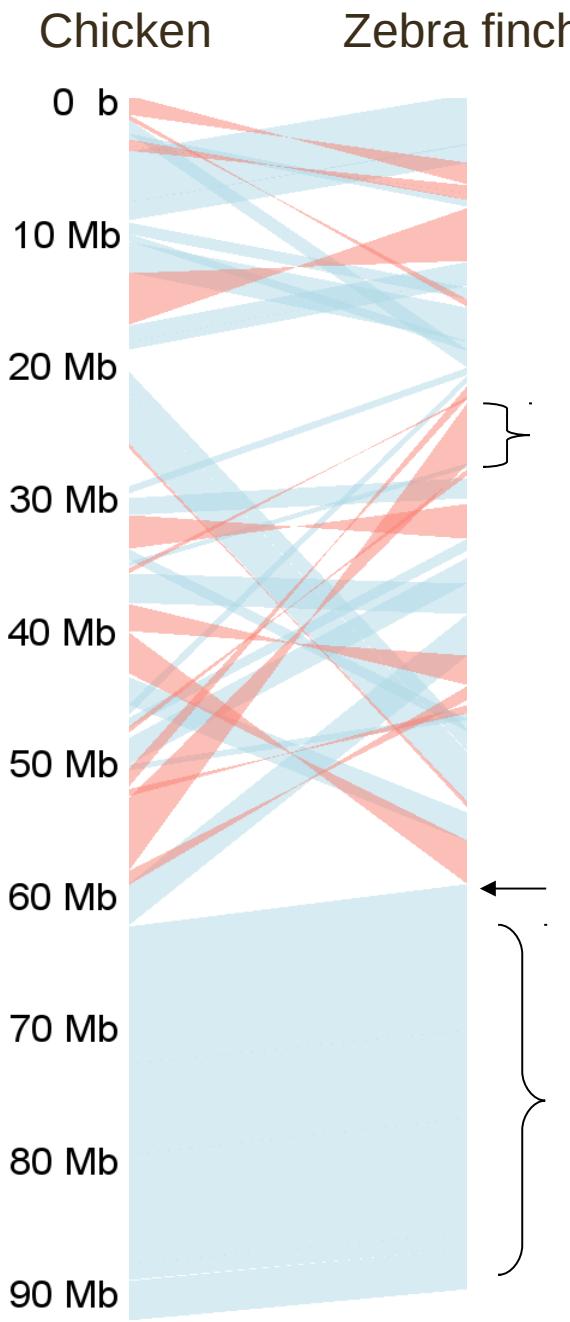
Copy number variants (CNVs)

CpG islands

Single Nucleotide Polymorphisms  
(SNPs)

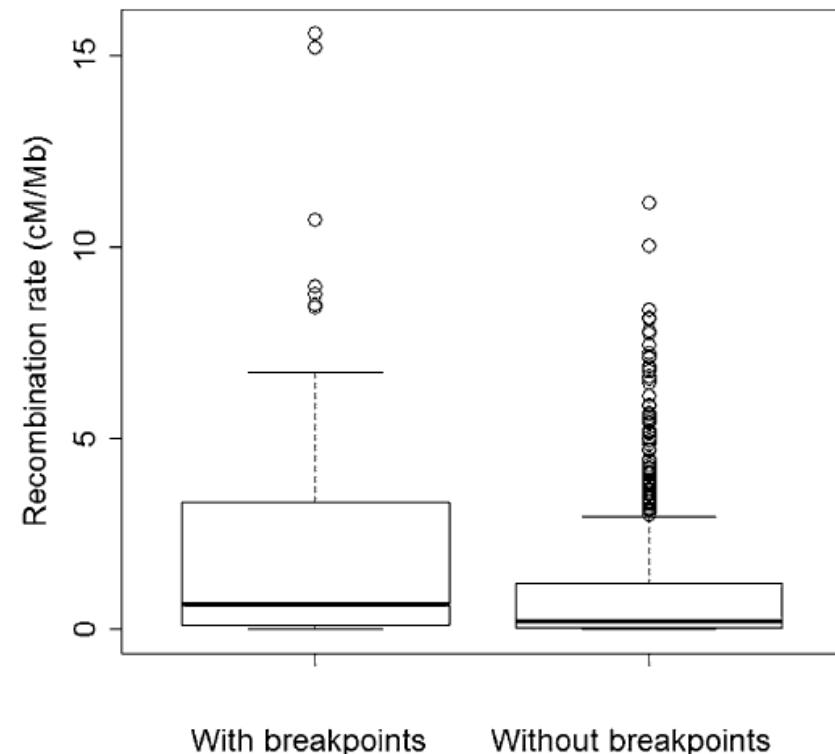


# Evolutionary Breakpoint Regions (EBRs)

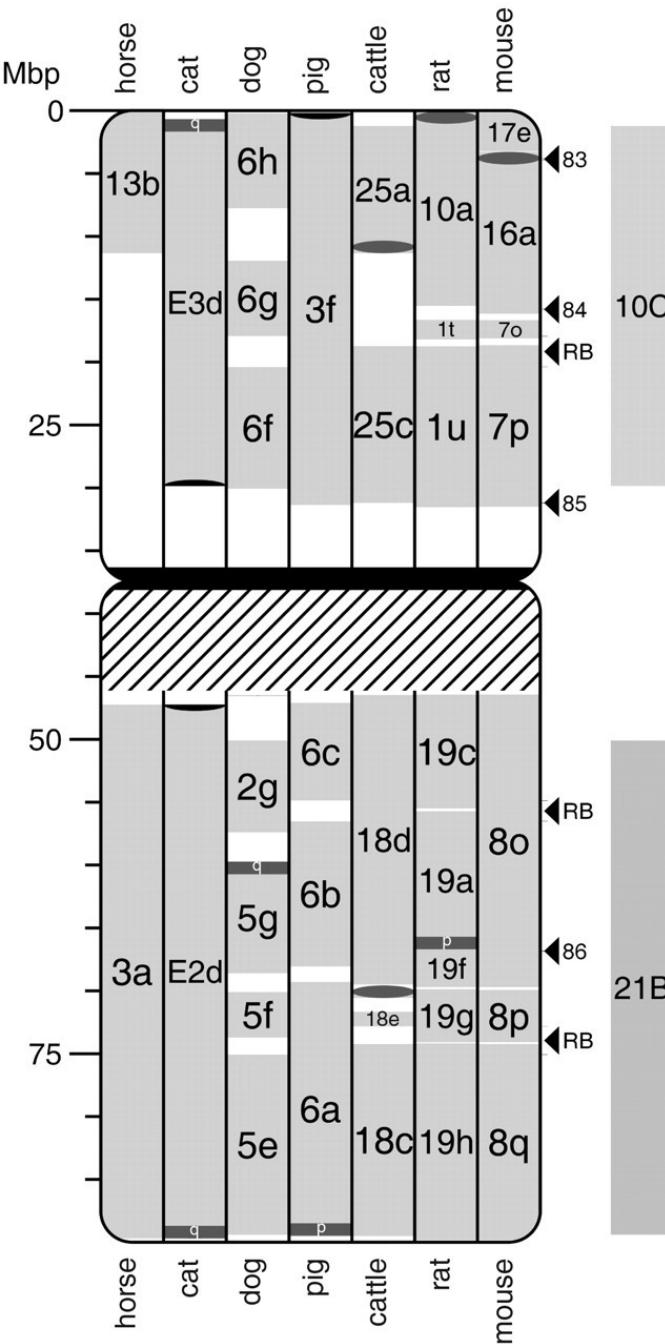


Comparison between chicken and zebra finch genomes

Association seen between evolutionary breakpoints, CNVs and recombination rates

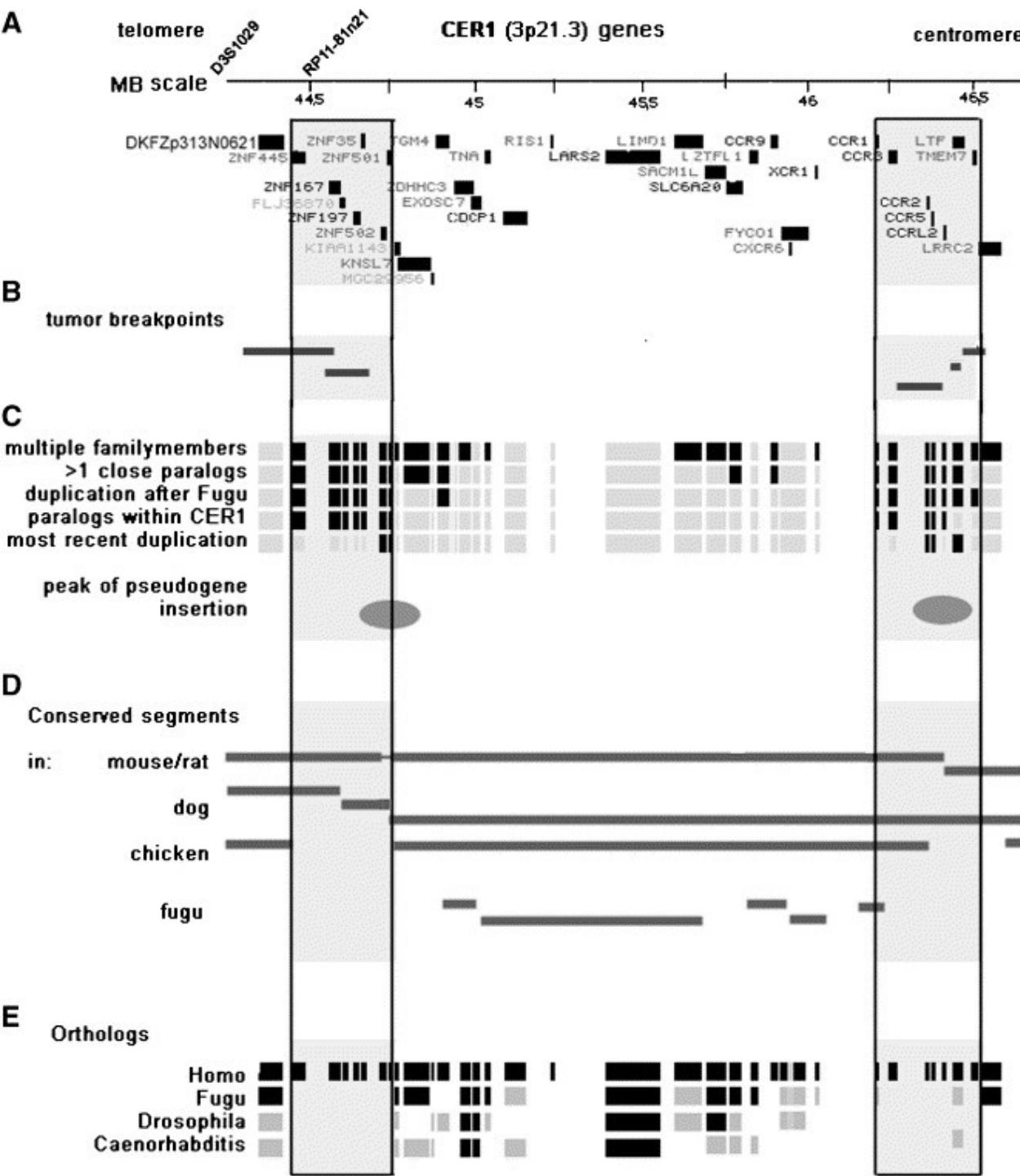


# Association of EBRs with cancer



Arrows mark human cancer breakpoints

# Association of EBRs with cancer

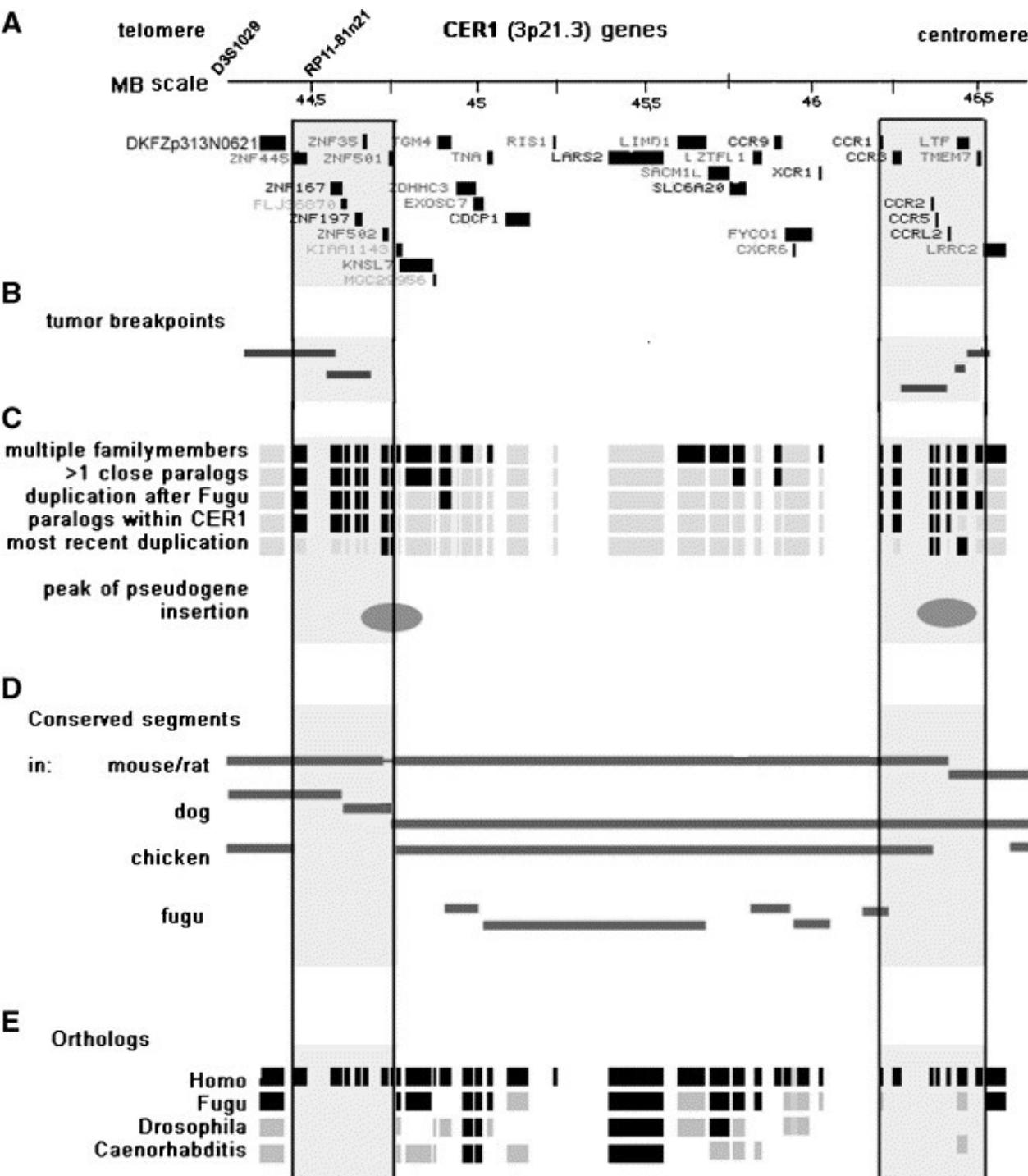


Comparative genomic analysis of CER1 border regions

2.4Mb Common Eliminated Region 1 from 3p21.3

Regularly lost in tumors derived from chromosome 3 microcell hybrids with mouse or human background

# Association of EBRs with cancer



Tumor breakpoint cluster regions characterized by:

- clustering of paralogs and high frequency of retroposed pseudogenes
- nonrandom occurrence of synteny breakpoints, (elevated evolutionary fragility of the border regions)

# Homologous Synteny Blocks (HSBs)

As well as conservation of breakpoint regions, also conservation of syntenic regions

- homologous synteny blocks between species

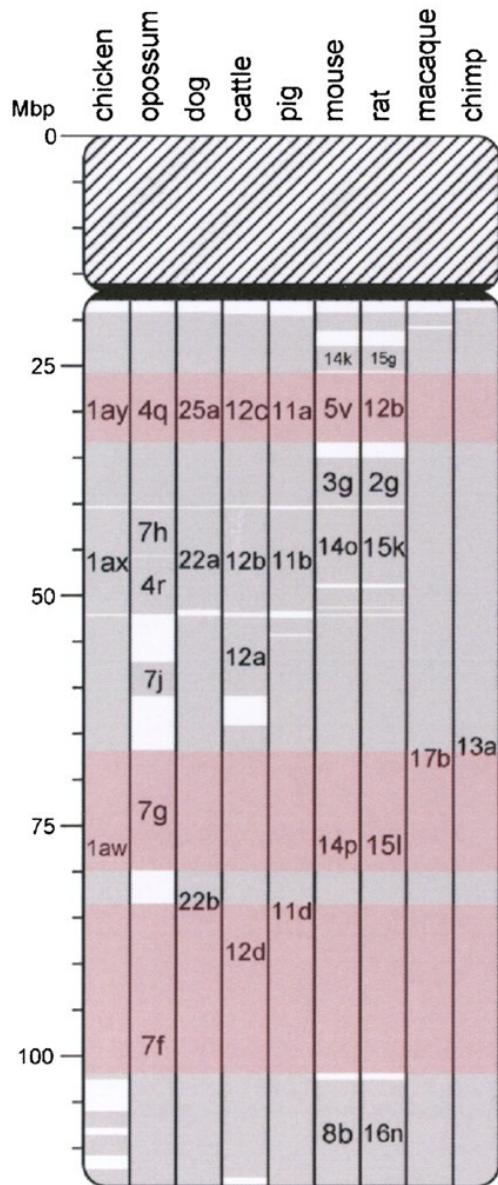
Detectable in comparisons between species as far diverged as human and chicken (LCA ~310Mya)

- ancient conserved segments of vertebrate genomes

# Homologous Synteny Blocks (HSBs)

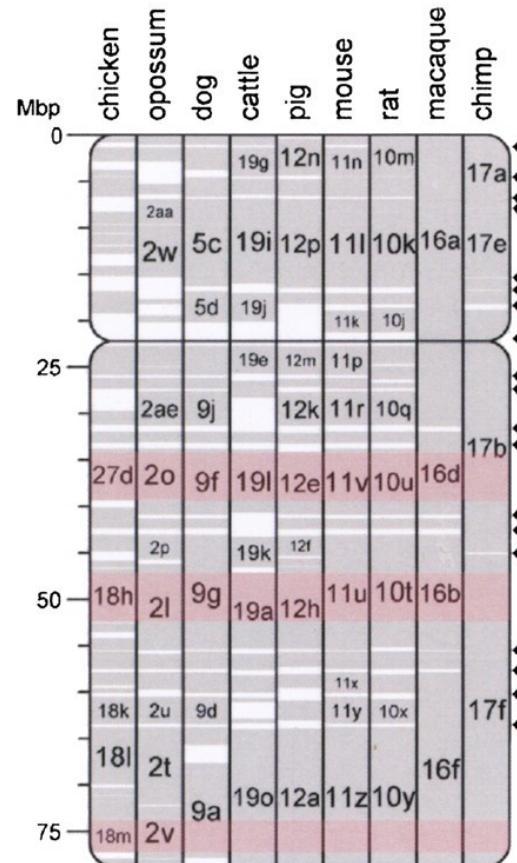
A

HSA13



B

HSA 17



structural variations  
segmental duplications  
CpG islands  
SNPs

# Homologous Synteny Blocks (HSBs)

Enriched for ancient conserved genes e.g.:

Notch- and Wnt-signalling pathways

cadherins (tissue organization)

gene networks involved in the formation and maintenance of cell junctions

genes in progesterone signalling network - essential for oocyte maturation

Too many large HSBs intact after 1Gyr divergent evolution to be chance - largest HSBs contain some critical genes e.g. in one HSB:

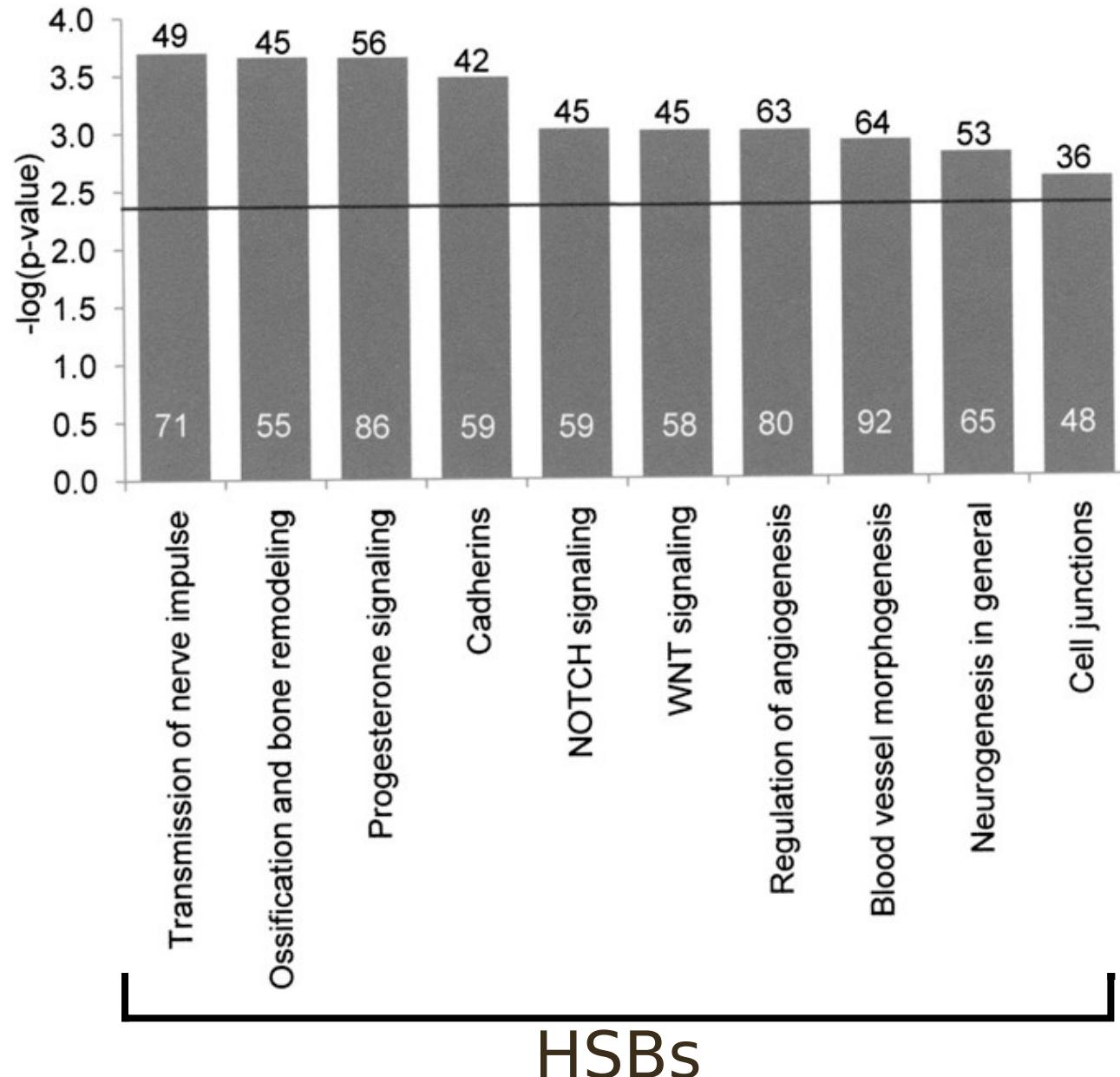
- HOXD gene cluster

- voltage gated sodium channels associated with brain and nervous system function

- two members of the DLX family of homeobox genes involved in craniofacial, limb, and bone development.

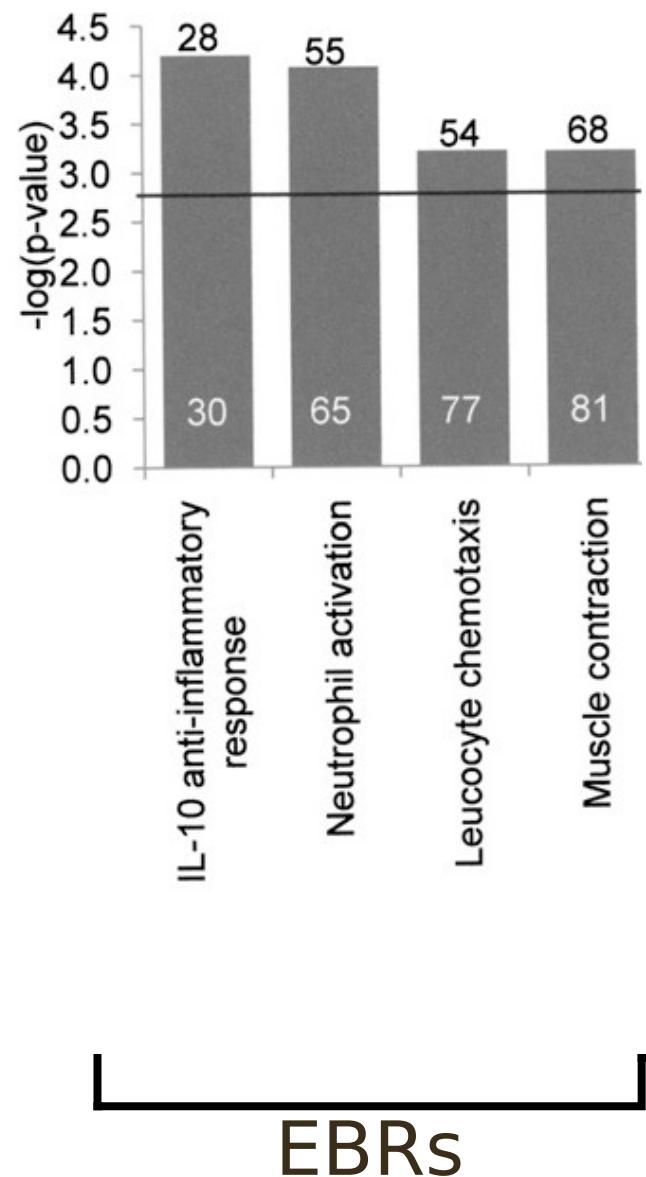
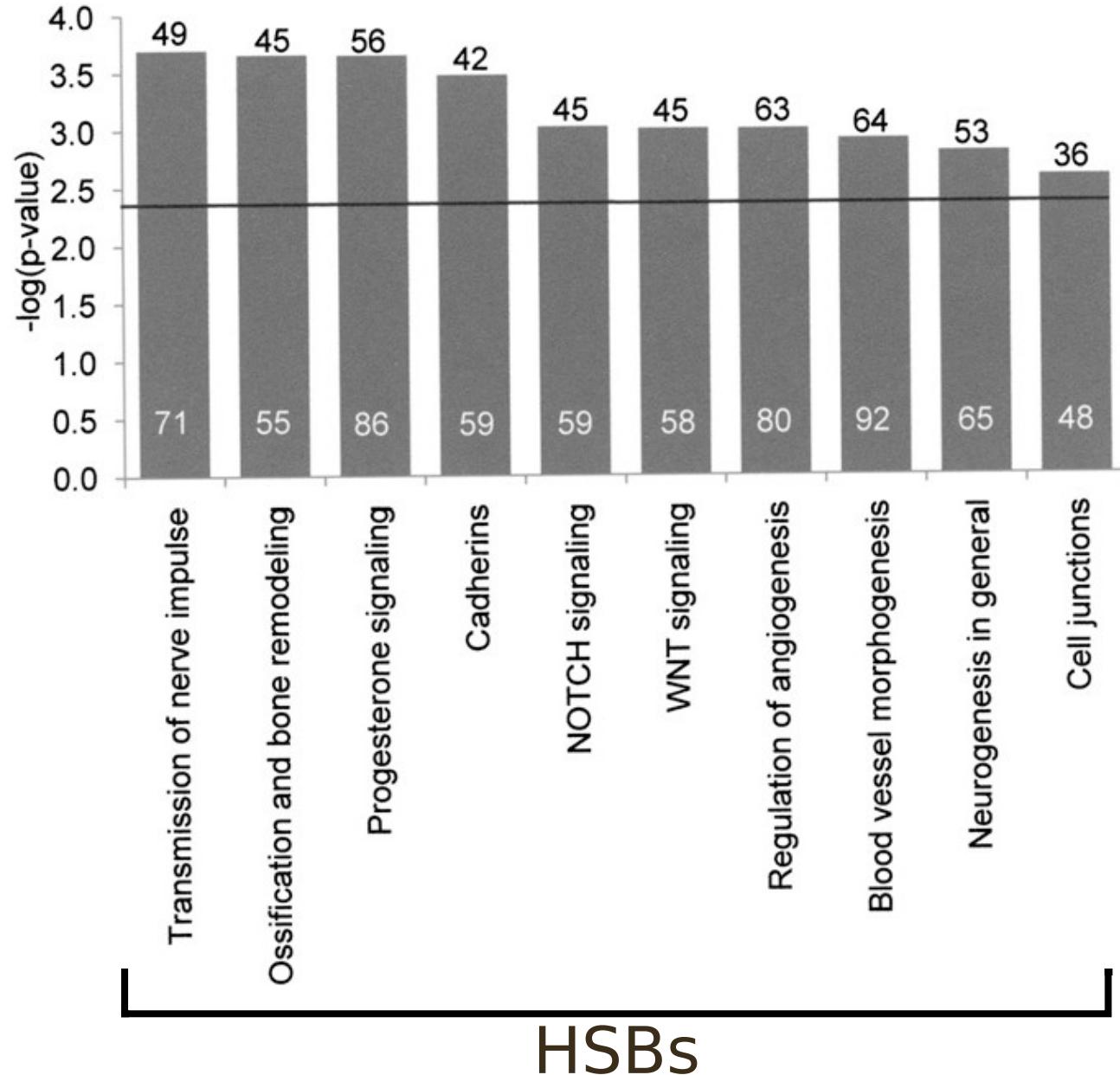
Breakages in these chromosomal regions may lower fitness

# Homologous Synteny Blocks (HSBs)



Preservation of important gene networks despite chromosomal rearrangement

# Homologous Synteny Blocks (HSBs)



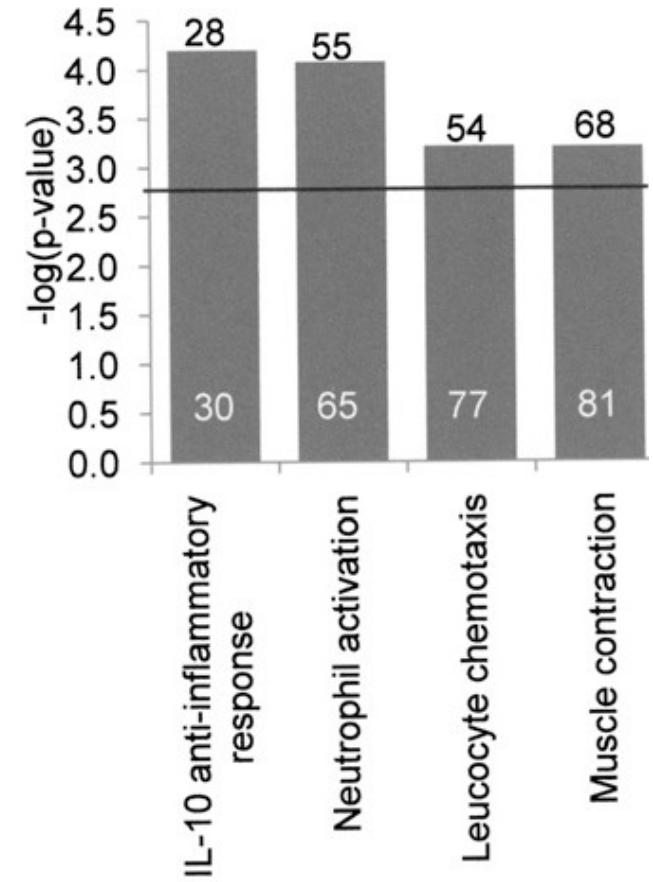
Preservation of important gene networks despite chromosomal rearrangement

# Genes within EBRs

EBRs are enriched for genes involved in response to external stimuli

Some evolutionary chromosome rearrangements may have adaptive value

Novel configurations of structural and regulatory loci involved in responses to environmental challenges



EBRs

# Nuclear organisation in evolution

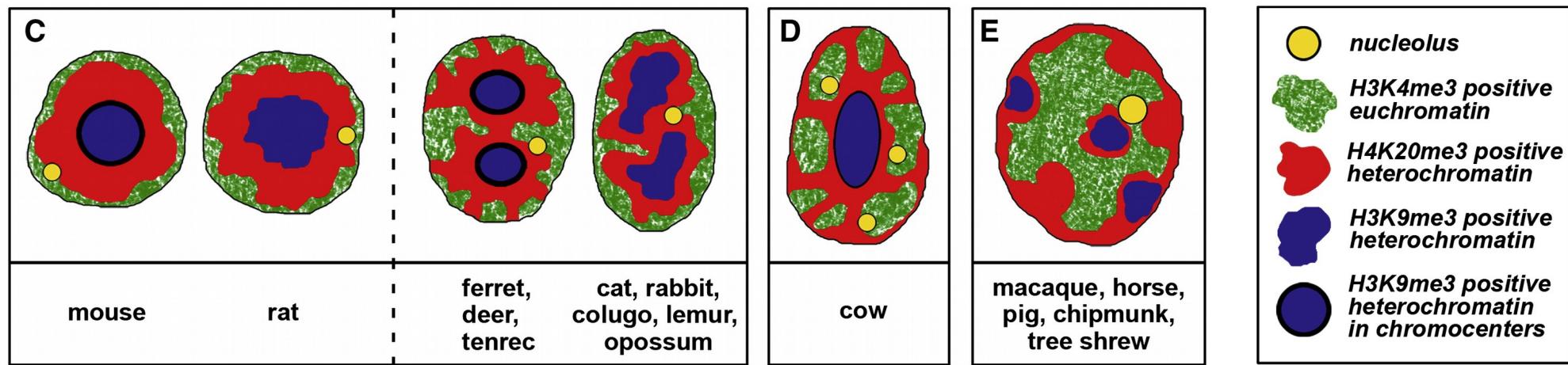
Position of chromosomes affects whether they will be able to interact

Evidence that 3D organisation of chromosomes (or syntenic regions) is conserved between mouse and human

For example, two loci distant in the human genome but adjacent in the mouse genome are significantly more often observed in close 3D proximity in the human nucleus than expected.

# Nuclear organisation in evolution

Retinal neuronal cells – organisation of heterochromatin in rod nuclei helps focus light in nocturnal mammals

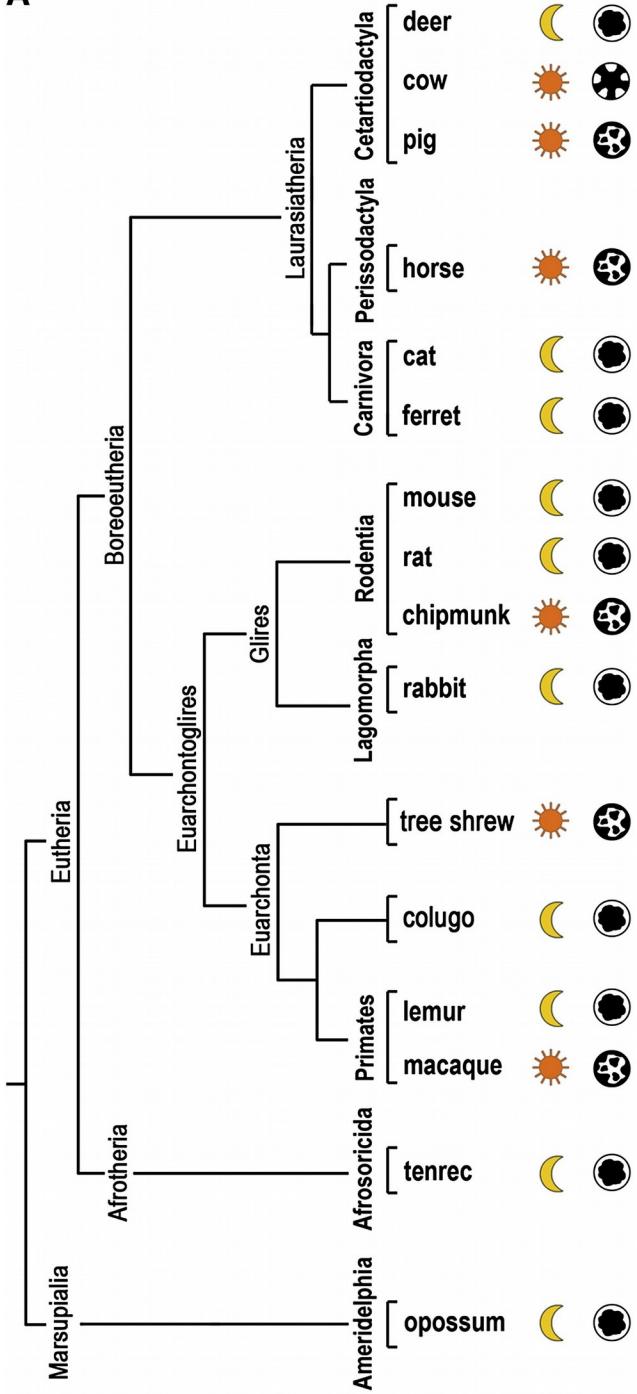


Inverted rod nuclei act as collecting lenses  
- channel light toward the light-sensing rod outer segments.

Conventional architecture prevails in eukaryotic nuclei because it results in more flexible chromosome arrangements, facilitating positional regulation of nuclear functions  
- e.g. gene silencing by migration to nuclear periphery

# Nuclear organisation in evolution

A



Inverted pattern is derived state

Evolved repeatedly throughout mammalian evolution

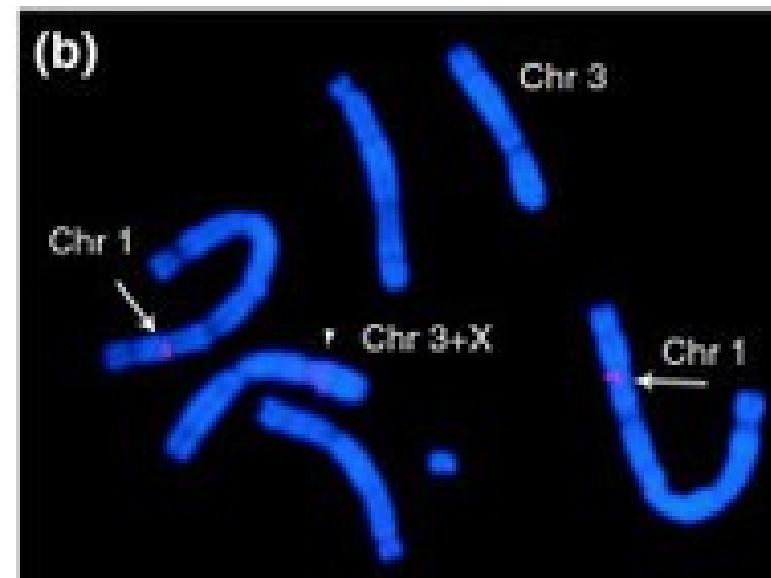
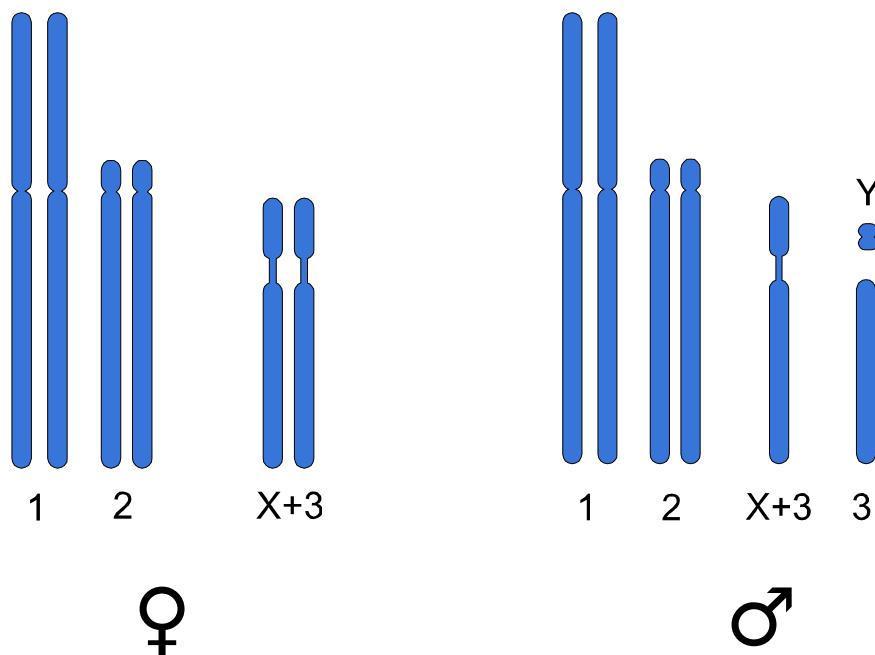
🌙 nocturnal / crepuscular activity (N, C)  
☀️ diurnal activity (D)

- inverted pattern
- conventional pattern
- intermediate pattern

# Extensive fusions in muntjac karyotypes

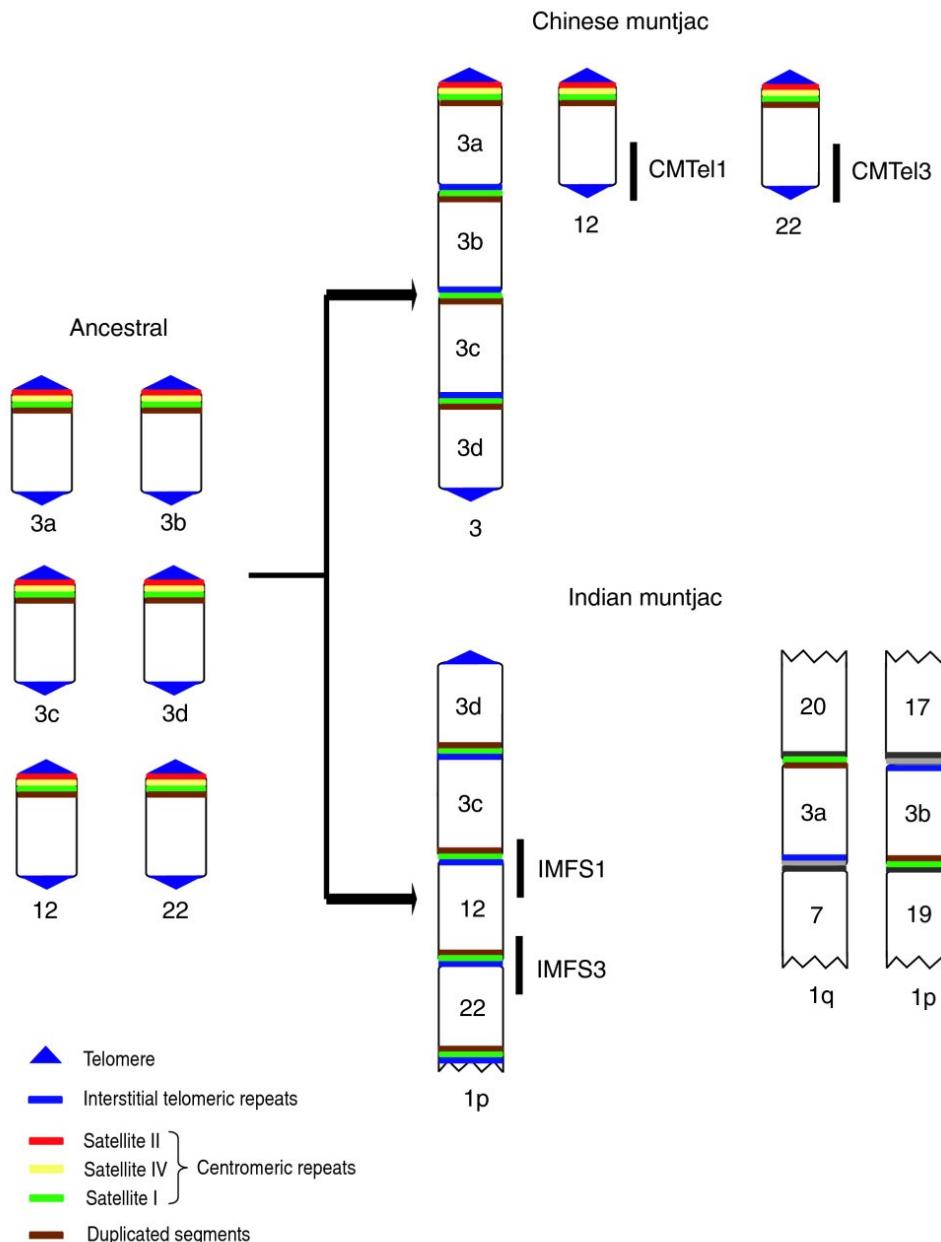
*Muntiacus muntjac*  
Indian muntjac deer

$$2n = 6 / 7$$



From Gen. Biol.  
(2008) 9:R155

# Extensive fusions in muntjac karyotypes



Reeves's muntjac  
(*Muntiacus reevesi*)

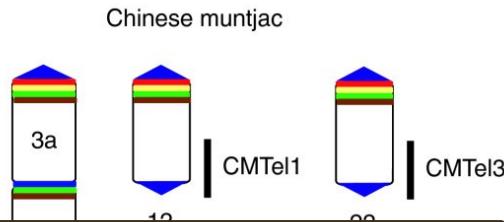
$2n = 46$



Indian muntjac  
*Muntiacus muntjac*

$2n = 6 \text{ ♀} / 7 \text{ ♂}$

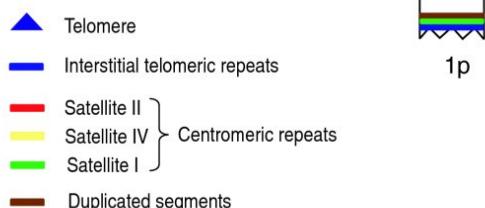
# Extensive fusions in muntjac karyotypes



Extensive fusions of ancestral chromosomes

Interstitial telomeres mark fusion points

Divergence about 3Mya – change can happen rapidly



Reeves's muntjac  
(*Muntiacus reevesi*)

$$2n = 46$$

1p



Indian muntjac  
*Muntiacus muntjac*

$$2n = 6 \text{ ♀} / 7 \text{ ♂}$$

# The target matters

Large changes to genome structure (e.g. karyotype) may have few actual consequences for gene regulation and expression

- Indian and Chinese muntjac can produce sterile hybrids

Conversely, small changes can have profound phenotypic effects

# The Chicken Polydactyly Locus



Ancona (Houdini)

4 toes



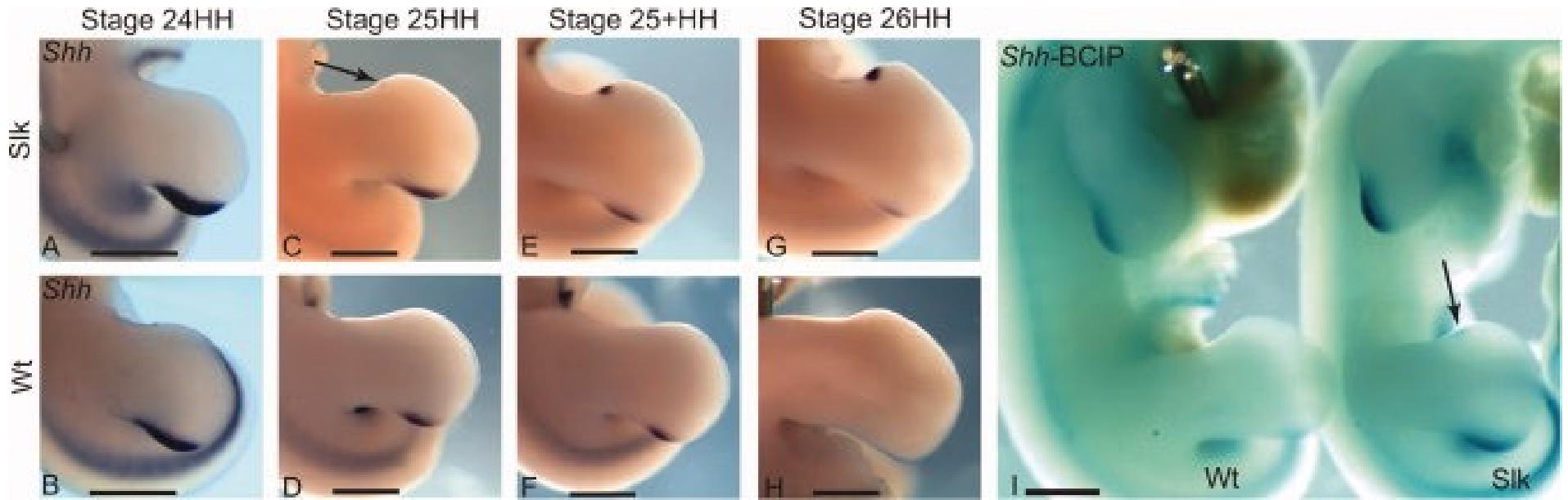
Silkie (A chicken with no name)

5 toes visible

Can have 6 or 7 toes per foot



# The Chicken Polydactyly Locus



Ectopic expression of Sonic hedgehog (*Shh*) in Silkies

-> causes excess digit 2 formation

Single SNP in the gene *LMBR1* ( a regulator of *Shh*)

C -> A transition

# Summary

Recombination, CNVs and structural chromosomal rearrangements are all linked

The same processes that underlie genome evolution can also underlie disease processes

Small genetic changes can have profound phenotypic effects

Interactions matter!

# Further Reading